2015

Individual Differences in the Use of Remote Vision Stereoscopic Displays

Marc Winterbottom
Wright State University

Follow this and additional works at: http://corescholar.libraries.wright.edu/etd_all

Part of the Industrial and Organizational Psychology Commons

Repository Citation
Individual differences in the use of remote vision stereoscopic displays

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

BY

MARC D. WINTERBOTTOM

B.A., Purdue University, 1996

M.S., Wright State University, 2000

2015

Wright State University
WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

April 30, 2015


____________________________________
Allen L. Nagy, Ph.D.
Dissertation Director

____________________________________
Scott N.J. Watamaniuk, Ph.D.
Graduate Program Director

____________________________________
Debra Steele-Johnson, Ph.D.
Chair, Department of Psychology

____________________________________
Robert E.W. Fyffe, Ph.D.
Vice President for Research and Dean, Graduate School

Committee on Final Examination

____________________________________
Allen L. Nagy, Ph.D. – Wright State University

____________________________________
Scott N.J. Watamaniuk, Ph.D. – Wright State University

____________________________________
James P. Gaska, Ph.D. – USAF School of Aerospace Medicine

____________________________________
Joe Houpt, Ph.D. – Wright State University
ABSTRACT

Winterbottom, Marc D., Ph.D., Department of Psychology, Human Factors Psychology Program, Wright State University, 2014. Individual differences in the use of remote vision stereoscopic displays.

With the introduction of the next generation of aerial refueling tankers, such as the KC-46, boom operators will use relatively recently developed indirect view stereo displays in place of direct view crew stations. Existing vision standards for boom operators were developed during the 1950s and may not be adequate for medical screening for KC-46 boom operators. Mild anomalies in binocular alignment, currently allowed by USAF vision standards, may permit stereopsis, but may also predispose those individuals to visual complaints such as eye-strain or headaches when viewing stereoscopic displays.

The purpose of this research was to measure individual differences in performance with the use of a simulated remote vision system (RVS) during a simulated aerial refueling task; and to evaluate the relationship between individual differences in refueling performance and individual measures of quality of vision. To accomplish this research, a simulated RVS aerial refueling crew station was developed based on specifications provided by the USAF KC-46 Program Office and The Boeing Company. Experiment 1 was designed to simulate a “fighter drag” operational scenario where a boom operator repeatedly refuels receiver aircraft. Twenty-seven participants with varying quality of vision were recruited for Experiment 1. Each participant’s vision was tested using the existing USAF test battery and a battery of newly-developed computer-based vision tests. In Experiment 2, the same RVS simulation was used to evaluate the overall effect of stereo viewing condition on refueling performance (2D, normal stereo, and hyper-stereo).

The results of Experiment 1 reveal that refueling performance and level of discomfort are clearly dependent on quality of vision. Although most participants were generally comfortable using the simulated RVS, a few participants reported high levels of discomfort. Two vision tests were highly correlated with aerial refueling performance: minimum contrast sensitivity and fusion
range. For young observers, minimum contrast sensitivity was also highly correlated with reported discomfort. Most of the standard USAF vision tests were not correlated with either RVS refueling performance or reported discomfort. However, vertical phoria was significantly correlated with both performance and comfort. The results of Experiment 2 reveal that the introduction of stereo and hyper-stereo improved RVS refueling performance.

*Keywords:* remote vision system, aerial refueling, stereoscopic display, 3D displays, depth perception, medical standards, vision standards, Flying Class III vision standards, depth perception standards
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. GENERAL METHODS</td>
<td>37</td>
</tr>
<tr>
<td>III. EXPERIMENT ONE – METHOD</td>
<td>52</td>
</tr>
<tr>
<td>IV. EXPERIMENT ONE – RESULTS AND DISCUSSION</td>
<td>58</td>
</tr>
<tr>
<td>V. EXPERIMENT TWO – EFFECT OF STEREO ON REFUELING PERFORMANCE</td>
<td>86</td>
</tr>
<tr>
<td>VI. EXPERIMENT TWO – RESULTS</td>
<td>88</td>
</tr>
<tr>
<td>VII. GENERAL RESULTS AND DISCUSSION</td>
<td>92</td>
</tr>
<tr>
<td>VIII. CONCLUSIONS AND FUTURE WORK</td>
<td>104</td>
</tr>
<tr>
<td>IX. REFERENCES</td>
<td>110</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1-1. Left. Zone of comfort described by Shibata et al. (2011) relative to ZCSBV and Percival's zone of comfort. Units are in diopters (D). Right. ZCSBV and phoria measured according to procedure described in this research. This participant exhibits a typical phoria line – exophoric at near and esophoric at far. ................................................................. 9

Figure 1-2. Visual performance/Flight Safety aerial refueling simulation (Lloyd & Nigus, 2012). Air Force contract N61339-03-D-300. ........................................................................................................ 14

Figure 1-3. Illustration of how subpixel shift in left and right eye images can be accomplished using antialiasing/blurring. Enlarged images of the left and right eye stereo acuity stimuli described in the methods below (left). Grayscale values of one row of pixels from the center portion of the stereo ring for each eye (right). ....................................................................................... 15

Figure 1-4. Threshold disparity as a function of pixel pitch and antialiasing filter kernel (half-max) width. From Lloyd (2013) presentation at the Stereoscopic Displays and Applications Conference. Used with permission of the author. ........................................................................................................ 16

Figure 1-5. Image distortion in a hyper-stereoscopic display system for an arbitrary camera configuration. The figures on the left show the position of a circular object relative to a pair of cameras positioned at 0 mm. The size of the object is constant, but its position in depth changes from 600 mm in the top left figure to 2000 mm in the lower left figure. The corresponding figures on the right illustrate the distortions that can occur with hyper-stereoscopic displays. The dashed line represents the display surface. Note that imagery behind the display is compressed while imagery in front of the display is elongated (computed based on Woods et al., 1993). ............... 19

Figure 1-6. Vertical misalignment for a hyper-stereoscopic camera configuration (computed based on equations described in Woods et al., 1993). The red and blue dots represent the horizontal and vertical offset of points on a surface viewed through hyper-stereo cameras on a 3D display that is approximately 600 mm wide x 400 mm tall. The red and blue dots represent the right vs. left camera image positions........................................................................................................ 20
Figure 1-7. Example of dipvergence in arc minutes as a function of horizontal image position as displayed on screen approximately 600 mm wide. ................................................................. 21

Figure 1-8. Howard-Dolman (left), AO Vectograph (top right), Titmus test (bottom right) .......... 25

Figure 1-9. Stereo acuity thresholds from Howard (1919) and from Zaroff et al. (2003) .......... 30

Figure 1-9. Howard's (1919) stereo acuity data in comparison to current USAF standards .......... 32

Figure 2-1. Simulated remote vision system aerial refueling task apparatus .......................... 38

Figure 2-2. Phoropter used to administer phoria and fusion range tests .............................. 39

Figure 2-3. Logitech game controller and Saitek joystick used to enter responses for computer-based vision tests ................................................................. 40

Figure 2-4. Stereo-Optical Armed Forces Vision Tester ....................................................... 41

Figure 2-5. Nvidia 3D Vision active shutter glasses (left) and passive polarizing 3D glasses supplied with the ViewSonic 3D monitor (right) ............................................................. 43

Figure 2-6. Near stereo test stimulus .................................................................................... 45

Figure 2-7. Schematic view of stereo acuity rings. Left: small ring in front of larger ring; center: small ring behind larger ring, right: eye indicating relative viewing position ............................. 46

Figure 2-8. Vertical fusion range stimuli (showing both left and right eye images) ................. 47

Figure 2-9. Horizontal fusion range stimuli (left and right eye images) .................................. 48

Figure 2-10. Band-pass filtered Landolt C stimuli. Figure provided by AST. Image used with permission .................................................................................................................. 49

Figure 2-11. View from the KC-10 ARO crew station ........................................................... 50

Figure 2-12. Questionnaire items adapted from Shibata et al. (2011) ................................. 51

Figure 3-1. Images of the aspect angle task target aircraft as displayed on the panoramic displays at different distances .......................................................................................... 53

Figure 3-2. Images of the aspect angle task target aircraft as displayed on the 3D display (boom camera view) at different distances .............................................................................. 53
**Figure 3-3.** Receiver aircraft positions viewed through the simulated RVS. Pre-contact position (top left), contact position 1 (top right), contact position 4 (lower left), and contact position 5 (lower right)........................................................................................................................................56

**Figure 3-4.** Approximately 30 seconds of data recorded at 24 Hz for one subject. In this example, nozzle distance to receptacle, boom pitch, boom yaw, and boom extension are shown.............57

**Figure 3-5.** Boom position data recorded at 24 Hz were reduced to a few metrics that were used to evaluate refueling performance. In this example, time to successful coupling (trial duration), number of attempts (number of trigger presses to initiate boom to receiver lock), and nozzle distance to receptacle at attempt are shown for an 80 minute time period for one subject...........58

**Figure 4-1.** Average values for focus/vergence difficulty, viewing comfort, and eye-tiredness throughout 2-hour refueling task ........................................................................................................................................59

**Figure 4-2.** Histogram showing the number of subjects reporting different levels of eye-tiredness at the conclusion at the end of the 2-hour refueling task. ........................................................................60

**Figure 4-3.** AFVT/AO Vectograph stereo acuity frequency histogram.............................................61

**Figure 4-4.** Frequency histograms for horizontal far (top) and near (bottom) phorias. ..........61

**Figure 4-5.** ZCSBV for two subjects........................................................................................................62

**Figure 4-6.** ZCSBV area frequency histogram. ..................................................................................62

**Figure 4-7.** Near and far stereo acuity histograms. .................................................................................63

**Figure 4-8.** Relationship between USAF standard stereo acuity tests and experimental computer-based stereo acuity test.........................................................................................................................64

**Figure 4-9.** Relationship between experimental near and far stereo acuity tests. .........................64

**Figure 4-10.** Horizontal and vertical fusion ranges for two subjects. Fusion break points are shown in blue, fusion recovery points are shown in red (log arc minutes). ...........................................65

**Figure 4-11.** Relationship between log horizontal fusion range and log ZCSBV area.............66

**Figure 4-12.** Contrast sensitivity (CS) test results for two subjects. CS functions are shown for binocular (OU), left eye only (OS), and right eye only (OD).........................................................................................66
Figure 4-13. Frequency histogram showing the range of AULCSF scores (OU) .......................67

Figure 4-14. Average aircraft aspect angle recognition range for each viewing condition ..........68

Figure 4-15. Relationship between AST visual acuity and aspect angle (AA) recognition range for the center panoramic display. ........................................................................................................69

Figure 4-16. The range of values obtained for Number of Contacts per block (top left), Number of Collisions per block (top right), average Number of Attempts (middle left), Mean Duration (middle right), Mean Distance (bottom left), and composite aerial refueling (AR) performance (bottom right). ........................................................................................................71

Figure 4-17. The relationship between OVT/AOV stereo (top left), far phoria (top right), near phoria (bottom left), and vertical phoria (bottom right) and composite AR performance. The horizontal dashed lines indicate 1 standard deviation below the mean for aerial refueling performance and the vertical solid lines indicate the current FCIII minimum vision standard for each test vision test. ........................................................................................................73

Figure 4-18. The relationship between AFVT/AOV stereo (top left), far phoria (top right), near phoria (bottom left), and vertical phoria (bottom right) and average discomfort ratings. ...........74

Figure 4-19. The relationship between FCIII status and AR performance. One standard deviation below the mean for AR success rate is indicated by the dashed line. .........................75

Figure 4-20. The relationship between log ZCSBV and AR Performance (left), and to average discomfort (right). ........................................................................................................76

Figure 4-21. The relationship between average stereo acuity (left), fusion range (right) and AR Performance. One standard deviation below the mean is shown for AR Performance (dashed horizontal line) and each vision metric (solid vertical lines). ..................................................77

Figure 4-22. The relationship between average stereo acuity (left), fusion range (right) and average discomfort ratings. ........................................................................................................77

Figure 4-23. The relationship between contrast sensitivity (left), minimum contrast sensitivity (right) and refueling performance. ..................................................................................78
Figure 4-24. The relationship between binocular contrast sensitivity (left) and minimum contrast sensitivity (right) and average discomfort ratings.................................78

Figure 4-25. Difference in discomfort ratings by age group..............................................................79

Figure 4-26. Relationship between minimum contrast sensitivity and average discomfort (left) and between minimum contrast sensitivity and aerial refueling performance for participants 30 and younger. ..................................................................................................................80

Figure 4-27. Relationship between minimum contrast sensitivity, discomfort ratings, and refueling performance for participants over 30 years of age........................................................................................................80

Figure 4-28. The relationship between near phoria and average discomfort (left) and between near phoria and aerial refueling performance (right) for participants 30 and younger.................81

Figure 4-29. Relationship between near phoria, discomfort, and refueling performance for participants over the age of 30.................................................................81

Figure 4-30. The relationship between average discomfort ratings and aerial refueling performance. Results for participants over 30 years of age is shown on the left, results for participants 30 and younger is shown on the right.................................................................82

Figure 4-31. The relationship between the combined vision metric (CVM) and aerial refueling performance. One standard deviation below the mean aerial refueling performance is indicated with the horizontal dashed line. The solid vertical line represents 1 standard deviation below the CVM mean. The dashed vertical line indicates 2 standard deviations below the CVM mean. ..............83

Figure 6-1. Number of contacts/minute for each viewing condition................................................88

Figure 6-2. Aerial refueling performance for participants categorized as either “good” or “poor” based on their combined quality of vision scores.........................................................89

Figure 6-3. Aerial refueling performance for each participant in each viewing condition. Participants with combined vision metric scores above the mean are color coded in green while participants with combined vision metric scores below the mean are color coded in red. .........91
Figure 7-1. The relationship between contrast sensitivity (upper left), minimum contrast sensitivity (lower left), visual acuity (upper right), minimum visual acuity (lower right), and stereo acuity.
LIST OF TABLES

Table 1-1. USAF, USN, and USA Vision Standards. FCIII (scanners) are applicable to aerial refueling operators. ........................................................................................................35

Table 2-1. Luminance measurements obtained without glasses for the stereo acuity and fusion range tests. .................................................................................................................................................42

Table 3-1. Luminance and contrast measurements for the aspect angle aircraft for two displays and several distances........................................................................................................54

Table 4-1. Correlations among vision tests. .................................................................................................68

Table 4-2. Sensitivity and specificity values for each of the standard USAF vision tests. ........73

Table 4-3. Correlations between USAF standard vision tests and AR performance (left), and between experimental vision tests and AR performance (right). Tests resulting in a significant correlation (p < 0.01) are highlighted in green. .........................................................................................................................84

Table 4-4. Sensitivity and specificity for USAF standard vision tests and experimental vision tests (right). Tests resulting in a high level of both sensitivity and specificity (> 0.9) are highlighted in green........................................................................................................................................84

Table 4-5. Correlations between each vision test and average discomfort ratings..................85
ACKNOWLEDGMENTS

I greatly appreciate the guidance and support provided by my dissertation committee members: Dr. James Gaska (USAFSAM), Dr. Allen Nagy (WSU), Dr. Scott Watamaniuk (WSU), and Dr. Joe Houpt (WSU). I would especially like to thank Dr. James Gaska for all his guidance and support in developing what is in reality a whole program of research. I would also like to thank Dr. Charles Lloyd for his advice and support, without which this research could not have been completed.

Thanks for the support from the USAF School of Aerospace Medicine Aerospace Ophthalmology Branch in the 711th Human Performance Wing. In particular I would like to thank Col John Gooch and Dr. Steven Hadley for their support over the past several years. Thanks specifically to Alex van Atta for programming the extensive battery of vision tests and to Tommy Bullock for additional support with the aerial refueling simulation development. Thanks also to Dr. Steven Wright and Dr. Darrell Rousse for their guidance concerning USAF vision standards and taking time out of their busy schedules to administer a lengthy battery of vision tests that were a critical part of this research. Thanks to Kelly Bishop and Scott Humphrey for help with finding volunteers and many, many hours of data collection. And, thanks to Dr. James Rader and Dr. Rodger Vanderbeek, who made possible this program of research.

Thanks also to The Boeing Company and Pat Corbett who provided critical technical details; John Luu and Joe Rich at the KC-46 Program Office (AFLCMC) for their support of this research; Flight Safety International and Dan Myers for providing the Vital X image generator; and to Dr. Lu Lesmes and Adaptive Sensory Technology for providing the specially modified contrast sensitivity test used in this research.

Thanks also to the DoD SMART Scholarship Program, without which this would not have been possible, and to Dr. Byron Pierce and Dr. Morley Stone for supporting my DoD SMART scholarship.

Most of all, thank you to my wife, WeiRong, and my two sons, Allen and Max, for their love and support throughout what has been a challenging few years.

Marc D. Winterbottom
30 April 2015
I. INTRODUCTION

With the introduction of the next generation of aerial refueling tankers (e.g., Boeing KC-46 for USAF, Airbus KC-30 for RAAF, Boeing KC-10 for RNLAF, KC-767 for JASDF), in which aerial refueling operators (AROs), or boom operators, will use relatively recently developed indirect view stereo displays in place of direct view crew stations, existing vision standards for boom operators may not be adequate (Gooch, 2012; Konishi, 2012; Singh, 2012; Smart & Singh, 2012; Smith, 2012). In particular, the level of stereo acuity and oculomotor capabilities required to maintain stereo fusion with these new stereoscopic remote vision displays in rested and fatigued states are generally unknown. As Lambooij, Fortuin, Ijsselsteijn, Evans, & Heynderickx (2011) note, mild anomalies (currently allowed by USAF vision standards) in binocular alignment may permit stereopsis, but may also predispose those individuals to visual complaints such as asthenopia (eye-strain) or headaches. These visual complaints may not occur under normal viewing conditions, but may arise under more unnatural viewing conditions, such as viewing stereoscopic content. Although 3D displays have been in use for many years, their popularity has grown in recent years, and the sales of 3D displays for television, movies, and gaming has grown considerably. Additionally, head-mounted display (e.g. Joint Strike Fighter) and remote view display applications (e.g. tele-robotic surgery, remotely controlled ground vehicles, remote view aerial refueling) have drawn attention to the need for more research on the use of stereoscopic imagery. Stereoscopic displays offer a number of potential benefits (Kooi & Toet, 2004):

1. Aid in encoding large amounts of complex information (e.g. 3D modeling of complex structures).
2. Improve perceptual separation of important details, especially in noisy/complex scenes.
3. Signal can be separated from noise more easily in a 3D display.
4. Binocular “unmasking”. Detection thresholds in binocular noisy scenes are lower than in monocular scenes (e.g. targets in turbid water on underwater cameras, breaking camouflage).

However, there are also serious drawbacks associated with 3D displays. Inconsistent cues may cause discomfort/eyestrain and 3D perception may be inaccurate or totally disrupted. In fact, reports of serious discomfort are very common. A recent study (Solimini, 2013) found that 55% of respondents reported discomfort after viewing a 3D movie (compared to 14% for 2D). Potential sources of problems include binocular asymmetry, or differences in left and right image quality, due to optics and/or filters; perceptual inconsistencies (e.g. vergence-accommodation mismatch, motion parallax-convergence mismatch resulting from the depth plane differing from the image plane) and cross-talk, resulting from incomplete separation of left/right eye images. In fact, some commercially available 3D displays come with warnings that users could experience altered vision, lightheadedness, confusion, nausea, and even convulsions.

Kooi and Toet (2004) systematically examined potential sources of discomfort. They used imagery of an office scene obtained with a digital camera. The camera was shifted 60 mm to obtain a pair of stereo images. For hyper-stereo pairs, the camera was shifted either 120 or 240 mm. Two of the actual depths for objects in the scene were 29 cm and 55 cm in front of the back wall at 220 cm from the camera. This produced disparities of 14 and 31 arcmin, respectively (for the orthostereoscopic conditions). Twenty-four subjects, aged 17-58 years, all with a minimum of 60-arcsec stereo acuity participated in the evaluation. The display viewing distance was 185 cm (i.e. midway
between the two foreground objects). The image pairs were manipulated in Photoshop in various ways:

1. ½ deg cyclorotation of both images
2. 1 deg rotation of 1 image
3. Magnification of 1 image (1.5 and 2.5%)
4. Meridional magnification of 3% (astigmatic effect – vertical, horizontal, trapezoidal) of one image
5. Horizontal shift (2, 3 PD) of one image
6. Vertical shift (1, 2 PD) of one image
7. Cross-talk (i.e. leakage, or inadequate separation of left and right eye images: 5, 15, 25%)
8. Various filters: red/green (simulating anaglyph glasses), Gaussian blur of one eye’s image, luminance (25% difference in one eye), color depth reduced to 4 bits in one eye

Each observer evaluated the comfort of the stereo image in comparison to the 2D reference image. The reference was first presented for 3.4 seconds, followed by a blank interval, then the stereo image was presented for 5 seconds. Subjects rated discomfort on scale of 1 – 5.

The results revealed that image manipulations affecting mainly the outer edges of the image did not have a large effect on comfort (e.g. rotation), probably due to the fact that subjects tended to fixate the center of the image, where distortions were small. They also found that vertical shifts tended to produce more discomfort. Cross-talk, blurred imagery, and color asymmetry, in particular, were also found to produce high levels of discomfort. Kooi and Toet note that disparities of 35 arcmin are generally comfortable to view, but that disparities of 70 arcmin are too large to view comfortably, and their results tend to support that rule of thumb. Several of these distortions, in particular hyper-stereo,
vertical misalignment, and cross-talk, may be present in the remote vision system (RVS) for the aerial refueling task. However, viewing periods will be much longer, and so there is the potential for significant levels of discomfort. However, because experimentation with lengthy and repeated viewing is time consuming and difficult, the extent to which increased exposure duration when viewing stereoscopic displays affects discomfort is not well-known. Research examining reports of discomfort often use relatively short viewing periods but very large misalignment, disparities, or other distortions in order to more efficiently examine factors affecting stereoscopic 3D display performance (e.g. Kooi & Toet, 2004; Lambooij, Fortuin, Ijsselsteijn, Evans, & Heynderickx, 2010). Emoto, Niida, Okano, & Member (2005) comment that amusement park stereoscopic displays are viewed for relatively short periods of time (30 min or less) and are therefore “visually bearable”. They also note that most studies examining discomfort involve only very short viewing periods, and go on to state that long term viewing does impact visual fatigue but do not cite any specific work examining the effect of viewing duration. For their research, Emoto et al (2005) adopted a 45 minute duration viewing period to examine the effects of large and repeated changes in vergence on fatigue when viewing a stereoscopic display. Hoffman, Girshick, Akeley, & Banks (2008) and Shibata, Kim, Hoffman, & Banks (2011) also used lengthy viewing periods to examine the effect of vergence-accommodation mismatch on fatigue and discomfort when viewing stereoscopic displays. It stands to reason that lengthier viewing periods could lead to increased fatigue and discomfort, and Shibata et al. (2011) do, in fact, show that symptom severity increased over the course of the 20 minute viewing period they used in their experimentation.
Kooi & Toet (2004) also evaluated the relationship of subjective ratings of comfort to quality of vision through the use of optometric tests (TNO stereopsis test, visual acuity at 5m in each eye, phoria, eye dominance). They found that image distortions were related to visual acuity – only those subjects with good acuity could detect some distortions in order to experience discomfort. The lack of correlation with subjects’ best visual acuity indicated that binocular vision is determined by the eye with decreased acuity. This is confirmed by the correlation between stereo acuity score and visual acuity in the poorer eye (r = -0.96). However, this correlation is not evident in Howard’s (1919) data with a much larger number of subjects.

A potentially significant problem for the wider use of stereoscopic displays, particularly in applications where operational task performance depends on information conveyed through a 3D display, is that a significant portion of the population may have binocular vision anomalies that affect stereopsis. Lambooij, Fortuin, Ijsselsteijn, Evans, & Heynderickx (2010) note that approximately 20% of the population may have anomalies in binocular vision. Some anomalies prevent normal stereopsis, such as strabismus or amblyopia. Other anomalies permit stereo, but may predispose patients to visual complaints (i.e. asthenopia, or eye-strain). Amblyopia is often the result of strabismus, where information from one eye is suppressed due to the misalignment. A less severe form of amblyopia may result from anisometropia, or unequal refraction. Microtropia, or monofixation syndrome, may result in less than perfect binocular vision. However, microtropias are also intermittent, and may be difficult to diagnose. Heterophoria occurs when vergence is not quite consistent across the two eyes, but is usually asymptomatic (“compensated”), since fusional reserve can usually compensate for small misalignment.
Lambooij et al. (2010) evaluated 50 volunteers on their quality of ocular alignment and classified them as having either 1) Good Binocular Status (GBS), or 2) Moderate Binocular Status (MBS). The screening tests included: 1) visual acuity; 2) refractive error; 3) stereopsis (Randot); 4) fixation disparity (angular adjustment required to align nonius lines); 5) heterophoria (Maddox Rod test); 7) cover test; 8) convergent fusion range; 9) divergent fusion range; 10) near point of convergence; 11) accommodative amplitude; 12) accommodation response; 13) accommodation facility; 14) vergence facility; and 15) slit lamp microscope. Based on these screening tests, 3 subjects were excluded from further testing (based on exclusion criteria such as strabismus, stereo blindness, heterophoria, etc.), and the remaining 47 subjects were classified according to 1) criteria for decompensated heterophoria, or 2) measurements greater than 1 S.D. outside the mean for a given screening test, with 4 or more deviations from normal resulting in MBS classification. According to these criteria, 38 subjects were classified as GBS and 9 as MBS.

Subjects were then asked to read aloud passages from the Wilkins Rate of Reading Test (WRRT) for 60 seconds using either a 2D or a 3D display (crossed disparity only). In the stereo condition, the relative disparity between text and frame in 3D condition was 1.5 degrees, which is beyond the 1-degree limit typically rated as comfortable, in order to stress subjects’ visual systems. In this condition, the text appeared to float in front of observers at 133 cm, while the frame at 0 degrees, was at 300 cm. A questionnaire was also administered to assess comfort. The results reported by Lambooij et al. (2010) reveal that 8 subjects were not able to fuse the 3D stimulus, so only 32 GBS and 7 MBS subjects’ data were evaluated. Both the GBS and MBS reported significantly more eyestrain for the 3D condition. However, the MBS subjects tended to report a greater degree of eyestrain and
other symptoms compared to GBS subjects. Lambooij et al. (2010) also report a significant effect of viewing condition on reading speed. Additionally, a significant interaction also occurred, with MBS subjects performing significantly worse on 3D display relative to 2D. Lambooij et al. (2010) go on to suggest that the WRRT could be used to screen observers prior to usage of 3D displays.

**Vergence-Accommodation Mismatch**

Another factor contributing to discomfort is the decoupling of vergence and accommodation, or VA mismatch. VA mismatch has been suspected as a source of discomfort for many years (Wann, Rushton, & Mon-Williams, 1995). However, a definitive link between VA mismatch and performance and discomfort was not established until more recently. In a series of experiments, Hoffman, Girshick, Akeley, & Banks (2008) and Shibata, Kim, Hoffman, & Banks (2011) systematically tested this hypothesis using unique displays that allowed them to simultaneously vary vergence and accommodative distance. As Hoffman et al. (2008) note, accommodation and vergence are coupled, and vergence changes occur faster in combination with accommodative changes, and vice versa. Thus, it would be expected that demanding stereoscopic tasks could be completed more rapidly when stimuli are consistent in terms of vergence and accommodation compared to when they are not. Hoffman et al. examined the effect of decoupling vergence and accommodation on threshold reaction time, threshold spatial frequency, and subjective ratings of comfort. They used random dot patterns depicting a grating varying in depth according to a sine function. Their results revealed that subjects’ reaction time to discriminate grating orientation substantially increased as the difference between accommodation/vergence distance increased (i.e. it took longer and was more
difficult for subjects to fuse the imagery in order to perform the task). They also found that thresholds for spatial frequency increased as the difference between accommodation/vergence distance increased. Finally, subjects reported a significantly greater degree of discomfort when the accommodation and vergence cues were inconsistent. Their data provide definitive evidence that VA mismatch contributes to discomfort.

Shibata et al. (2010) took this line of inquiry a step further and examined not only VA mismatch, but also the binocular status of the observers. Although the notion that VA mismatch could play a role in the use of visual displays is relatively new, as Shibata et al. note, it has been studied extensively in optometry and ophthalmology while monitoring patients’ adjustments to optical correction. As cited by Shibata et al., the zone of clear single binocular vision (ZCSBV) was first recognized by Donders in 1864, and Percival (1892) was the first to realize the significance of ZCSBV for fitting spectacles. In fact, the middle 1/3 of the ZCSBV, considered comfortable to view, is named Percival’s zone of comfort. Sheard’s zone of comfort is similar, but parallels the individual’s phoria line rather than the demand line. Two concepts that may be relevant to stereo displays emerge from the early work by optometrists: 1) phorias, and 2) the zone of clear single binocular vision. Phoria is the vergence posture of the two eyes when stimuli are viewed monocularly (i.e. resting vergence). Phoria is driven by the neural coupling with accommodation during monocular viewing. With exophoria the eyes tend to under-converge, while for esophoria the eyes tend to over-converge. Normal individuals tend to be exophoric at near, esophoric at far. However, there are large individual differences in phoria. To measure ZCSBV, the examiner adjusts the vergence stimulus using prisms for different fixed focal stimuli. The
examiner finds the maximum convergence and divergence for which the patient sees a single, well-focused target. The ZCSBV usually parallels phoria measurements as shown in Figure 1.

![Figure 1-1](image-url)

**Figure 1-1.** Left. Zone of comfort described by Shibata et al. (2011) relative to ZCSBV and Percival's zone of comfort. Units are in diopters (D). Right. ZCSBV and phoria measured according to procedure described in this research. This participant exhibits a typical phoria line – exophoric at near and esophoric at far.

Shibata et al. found that not only did subjective ratings of discomfort increase with increased accommodation/vergence mismatch, but that individual phoria measurements were also correlated with comfort ratings. Their results suggest that the zone of comfort is +/- 0.5 diopters (D) relative to a line with slope of 1 for vergence (x-axis) and accommodation (y-axis) as shown in Figure 1. Their phoria and ZCSBV measurements also confirm the findings of other researchers in showing that both varied considerably across subjects. On average, subjects tended to be esophoric for long focal distances, and
exophoric for near focal distances. Shibata et al. note that their results were generally similar to Tait (1951), who tested a very large number of subjects (4,880), but only for two distances. However, their zone of comfort measures appear to be narrower than either Sheard’s or Percival’s zone of comfort.

Despite the variability of the comfort ratings data, the measures of phoria and zone of comfort were predictive of ratings of discomfort in many cases (significant r scores range from approximately 0.25 to 0.5). They also note that a 1.2 D crossed disparity display was relatively more uncomfortable to view at far rather than near, but close distance displays were generally more fatigue producing. Finally, these authors also conclude that disparity behind the screen (uncrossed) is more uncomfortable relative to disparity in front of the screen (crossed) for longer viewing distances. But, the opposite is true for near distances: conflict in front of the screen was more uncomfortable than behind. Thus, driving apart accommodation and vergence leads to greater discomfort using a stereoscopic display. Placing the display surface at a near distance but presenting a vergence target in depth at a far distance is uncomfortable, and similarly, placing the display surface at a far distance but then providing a near vergence target in depth is also uncomfortable. Although it seems plausible that esophoria and exophoria might result in different outcomes in terms of user discomfort depending on the direction of accommodation/vergence mismatch and the viewing distance, Shibata et al. used the absolute value of the phoria measure in their analysis, and so did not investigate this particular issue.

**Previous ARO/Boom Operator Research with Stereo Displays**

Lloyd and Nigus (2012) investigated aerial refueling boom operator training for KC-10 aircraft using a simulation. KC-10 boom operators view receiver aircraft directly
through a window in the rear of the aircraft. Thus, although the viewing conditions are different, the refueling simulation discussed by Lloyd and Nigus is similar to that described in the current research. As they discovered, existing ground-based boom operator trainers (BOTS) do not use stereo displays. This was due to two factors: 1) conventional knowledge held that stereo was not useful beyond a few meters. Previous studies using electronic displays (Tidwell, 1990; Yeh & Silverstein, 1990) found stereo acuity thresholds of ~ 140 arcsec (i.e. many times higher than reported for real objects). Thus previous experience with inadequate displays may have led to the conclusion that stereo cues would be ineffective for larger distances. 2) Previous implementations of stereo displays in BOTS had been rejected by aircrew.

Based on interviews with KC-10 boom operators (BO’s) and boom operator instructors (BOI’s), Lloyd and Nigus found that existing (non-stereo) BOTS did not provide depth cues adequate for judging depth of approaching aircraft or distance between boom nozzle and receptacle. Due to these shortcomings, course developers could not move much training to ground-based trainers. This finding is unfortunate for several reasons. First, with several aircraft involved, and a full crew required for each tanker, the cost of aerial refueling training is tens of thousands of dollars per hour. Second, coordination and availability of not just tanker crew, but also receiver aircraft is problematic. Finally, an effective BOT would reduce time required for training, and would allow training with receiver aircraft less frequently encountered in airborne training missions.

Lloyd and Nigus (2012) identified the following visual cues that could potentially be available to BO’s to better support ground-based BOTS:
1. Binocular disparity: Normal observers can reliably discriminate disparities of 3 – 10 arcsec (i.e. ~ 1 foot at 60 feet).

2. Shadows: Depends on position of the sun or aircraft lighting at night

3. Motion parallax due to head motion: Reliable cue in aircraft, but conflicting cue in BOT due to distance to screen (however, AROs tend to keep head in fixed position and do not seem to routinely use this cue in practice).

4. Familiar size.

5. Vergence/accommodation distance: at a 60-foot viewing distance, roughly the distance from the ARO to the receiver aircraft in contact position, these are probably not powerful cues. These cues could be conflicting in the BOT due to vergence-accommodation mismatch.

Another visual cue that was not noted explicitly by these authors is the potential use of framing, or position of aircraft relative to the windscreen or other reference points.

Lloyd and Nigus evaluated the effect of several display factors on the estimation of distance and subjective comfort ratings in a boom operator simulation (see Figure 1-2):

1. Stereo vs. non-stereo (for non-stereo imagery, the same apparatus used, but IPD set to zero)

2. Collimated vs. non-collimated (collimation uses optics to produce imagery at optical infinity, whereas the non-collimated display viewing distance was 1.1 m).

3. Head-tracking vs. none (i.e. motion parallax vs. none; head-tracking allowed the simulated image to move in conjunction with observer head movements, producing simulated motion parallax).

Their results revealed that 1) Distance estimates were significantly larger for non-stereo vs. stereo at contact distance, and that the standard deviation of the estimates were consistent with a stereo acuity of ~ 3-10 arcsec in the stereo viewing conditions; 2) Collimation resulted in smaller errors relative to non-collimated; and 3) there was a significant interaction between the stereo and head-tracking viewing conditions - tracking reduced performance on non-stereo displays, but had no effect on stereo display
performance. The authors concluded that the combination of collimation, stereo, and head-tracking resulted in the best performance in terms of both accuracy and comfort ratings. Their work provides some evidence for the utility of binocular disparity in the performance of an operational task, and also shows that currently available stereo displays can support stereo acuity-limited performance at distances beyond a few meters. The authors note that the relatively high-resolution displays (1.5 arcmin/pixel) and high quality antialiasing that were employed for their study may have been important factors. Of the cues to depth they identified in the list above, only a few of them will be reliable for the RVS: binocular disparity, shadows (depending on position, time of day, weather), and familiar size. However, binocular disparity, although magnified with the use of hyper-stereo, will vary depending on the receiver distance from the boom cameras (see discussion below). Familiar size may be less useful because the RVS will minify the receiver aircraft in the camera view, and will also distort the size and shape of the receiver aircraft (more discussion on this below). Vergence, accommodation, and motion parallax are not reliable cues due to the reasons described by Lloyd and Nigus, but also due to the fact that the RVS will introduce additional distortions. However, the position of the receiver aircraft within the display frame should continue to be a reliable cue.

As Lloyd and Nigus (2012) noted, previous research had indicated that stereo displays may have limited capability to accurately depict small disparities (e.g. Yeh & Silverstein, 1990). However, their results indicated that distance estimates were consistent with stereo acuity of approximately 5-10 arcsec. They speculated that the high-quality antialiasing with the Flight Safety IG combined with relatively fine display pixel pitch supported better depth discrimination than had previously been reported.
Lloyd (2012) further investigated the effect of both resolution and antialiasing. He noted that a system capable of producing “eye-limited” stereopsis cues requires the placement of lines and polygon edges with great precision. A typical simulation system today with ~ 2.2 arcmin pixels would require very good antialiasing to produce the small shifts in position necessary to support 3 – 10 arcsec disparity cues (see also Bach, Schmitt, Kromeier, & Kommerell, 2001). This is because shifting the left and right eye images based solely on the display pixel mosaic would result in shifts in depth much larger than the stereo thresholds of most observers (i.e. shifting the image by an entire pixel, or around 2 arcmin for a typical simulation display viewing distance, which is many times larger than a typical 10 arcsec stereo threshold). Figure 1-3 below illustrates how sub-pixel shifts can be obtained using antialiasing/blurring based on the stereo acuity stimuli used in the current research. As shown, the contrast profile of the stimuli shift slightly in opposite directions, 1/2 pixel in this example, rather than each by a whole pixel. This technique enables very
small disparities with the use of standard HD-resolution computer monitors and typical viewing distances.

![Image](image.png)

**Figure 1-3.** Illustration of how subpixel shift in left and right eye images can be accomplished using antialiasing/blurring. Enlarged images of the left and right eye stereo acuity stimuli described in the methods below (left). Grayscale values of one row of pixels from the center portion of the stereo ring for each eye (right).

Lloyd (2012) asked 4 observers (8 other observers failed a stereo acuity screening test, more on that later) to make in front of/behind judgments to obtain disparity thresholds while varying both resolution (pixel pitch: 0.48, 0.96, 1.44, 1.92, 2.4, 2.88 arcmin) and antialiasing level (varying the Gaussian edge blur for the target stimuli). He used a viewing distance of 60 ft. which is representative of viewing distances for KC-10 boom operators, and the target stimulus that was fixed in depth was sized to be similar to that of the refueling
receptacle on a receiver aircraft. The results show that for small pixels (0.48 arcmin) antialiasing made little difference. However, for larger pixels, weak or no antialiasing resulted in much larger thresholds – up to 25 arcsec as pixel size increased to 3 arcmin. A larger antialiasing kernel size decreased thresholds again up to a width of ~ 1 to 2 pixels. Thresholds rose again slightly for widths of 2 to 3 pixels (see Figure 1-4). Comfort ratings followed a similar pattern – with the highest comfort ratings found for small pixel sizes. Comfort decreased as pixel size increased, but moderate antialiasing ameliorated this effect to some degree. Lloyd (2012) found that mean thresholds under the best conditions were ~ 5.5 arcsec, which is consistent with estimates from previous work (McKee, 1983; Patterson & Martin, 1992). Thus, although previous work had suggested that stereo displays might only support disparities of 140 arcsec or larger, Lloyd’s (2012) results indicate that modern stereo displays can accurately depict disparities as small as a few arcsec provided that pixels are approximately 1.5 to 2 arcmin and that sufficient antialiasing is applied.

**Figure 1-4.** Threshold disparity as a function of pixel pitch and antialiasing filter kernel (half-max) width. From Lloyd (2013) presentation at the
Hyper-stereo

As Kooi and Toet (2004) found, the use of hyper-stereo can significantly reduce viewing comfort. However, a separate but equally important issue is the effect of hyper-stereo on performance. The use of hyper-stereo is becoming an important issue since it is being employed in a variety of applications, such as with HMDs and aerial refueling. Stuart et al. (2009) note that increased disparity results in distorted cues to distance/depth:

1. Increased convergence of observers eyes, which signals nearer distances
2. Disparity magnification leads to increased visual depth
3. Increased differential perspective between the 2 eyes reduces apparent distance

These lead to visual illusions that may affect flight performance (e.g. crater or bowl effect, distorted perception of slope, distorted time to contact estimates, etc.). However, it is not very well understood how observers might adapt to these distorted viewing conditions. Additionally, “adaptation” might be taken to mean different things.

1. Compensation: observers still experience distorted vision but adopt compensatory strategies
2. Recalibration: the relationship between visual cue and, in this case, depth is recalibrated (i.e. the mapping of visual cues onto motor responses is revised)
3. Reweighting: the discounting of the conflicting cue when computing a response (e.g., in this case weight monocular cues more heavily).

Stuart et al. (2009) asked three experienced pilots to fly five sorties with the Top Owl helmet-mounted night vision system (with intersensor differences of ~300 mm, or about 5x the typical IPD) over a 2 week period and asked them to perform several different
specific maneuvers. Interviews with the pilots revealed that the bowl effect was very noticeable in the first several sorties, but became less noticeable by the 5th sortie. Additionally, pilots did not report distortions in depth after landing and removing the helmet. One of the pilots, who had some previous experience with the Top Owl, had the same responses and perceived bowl effect on the first several sorties as the other two pilots who had no previous experience. Based on these observations, the authors suggest that rather than recalibration, pilots simply began to weight monocular cues more heavily to compensate for hyper-stereo distortions.

In a related series of experiments (Flanagan, Stuart, & Gibbs, 2007a, 2007b; Stuart, Flanagan, & Gibbs, 2007) the effect of hyper-stereo on time to contact estimates, distance estimation, and slope perception was examined. These authors found that hyper-stereo led to an underestimation of time to contact, and an increase in the perceived slope of a textured surface. However, the results were highly variable across individuals, and the effect of hyper-stereo distortions was less than predicted. It is important to note that these studies were intended to simulate the Top Owl camera separation, field of view, and convergence distances. Woods, Docherty and Koch (1993) describe the mathematical equations for determining the extent of the image distortions resulting from camera toe-in, convergence angle, display size, viewing distance and other characteristics of the camera and display system. Figure 1-5 below is based on these equations and illustrates the distortions created for a circular object for a camera separation of 75 mm, convergence distance of 0.9 m, viewing distance of 0.9 m, eye separation of 65 mm, and display width of 300 mm.
Figure 1-5. Image distortion in a hyper-stereoscopic display system for an arbitrary camera configuration. The figures on the left show the position of a circular object relative to a pair of cameras positioned at 0 mm. The size of the object is constant, but its position in depth changes from 600 mm in the top left figure to 2000 mm in the lower left figure. The corresponding figures on the right illustrate the distortions that can occur with hyper-stereoscopic displays. The dashed line represents the display surface. Note that imagery behind the display is compressed while imagery in front of the display is elongated (computed based on Woods et al., 1993).
Figure 1-6 illustrates vertical parallax (dipvergence) resulting from camera toe-in.

For simulation and training applications, a commonly used rule of thumb is that vertical disparity greater than 5 arcmin will be uncomfortable.

**Figure 1-6.** Vertical misalignment for a hyper-stereoscopic camera configuration (computed based on equations described in Woods et al., 1993). The red and blue dots represent the horizontal and vertical offset of points on a surface viewed through hyper-stereo cameras on a 3D display that is approximately 600 mm wide x 400 mm tall. The red and blue dots represent the right vs. left camera image positions.
In this example, the camera toe-in results in a vertical parallax in the corners of the display of over 5 mm. The vertical disparity in this case would be approximately 22 arcmin, greatly exceeding the value considered acceptable. Figure 1-7 shows how vertical misalignment varies with horizontal position.

![Graph showing dipvergence](image)

**Figure 1-7.** Example of dipvergence in arc minutes as a function of horizontal image position as displayed on screen approximately 600 mm wide.

As the work by Woods et al. (1993) illustrates, a hyper-stereoscopic display system can result in depth plane curvature, depth non-linearity, shear distortion, and keystone distortion and vertical parallax (dipvergence). Banks, Read, Allison, and Watt (2012) also caution that camera toe-in can lead to vertical disparity, and may also reduce perceived depth. They go on to note that little work has been done to examine depth distortions caused by camera toe-in. A key aspect of the research presented here is to replicate to the greatest extent feasible, any of these distortions that may result from the particular
configuration of the KC-46 RVS. The characteristics of the RVS were simulated based on
detailed and proprietary data provided by the Boeing Co.

**Stereo Acuity**

Stereo acuity thresholds are comparable to other forms of hyper-acute, and can be
as small as approximately 5 – 10 arcsec (McKee, 1983; Lloyd & Nigus, 2012; Patterson,
1992). However, various factors can affect stereo acuity thresholds, such as: luminance,
contrast, spatial frequency, eccentricity, vergence position, and motion/temporal
frequency. For best stereo acuity, the target should be near the plane of fixation (however,
see also Zaroff, Knutelska, & Frumkes, 2003, discussed below), and the target and
reference should be 10 to 30 arcmin apart (McKee, 1983). Legge & Gu (1989) showed
that increasing contrast resulted in gradual improvement in threshold disparity (front/back
discrimination for sine wave pattern relative to a reference), with the slope approximately
constant across spatial frequencies ranging from 0.5 to 3.5 cpd. Interestingly, Legge and
Gu found that disparity thresholds for their 7 subjects were highly correlated with each
subject’s contrast sensitivity at the same spatial frequency (r = 0.84). Legge and Gu also
found that disparity thresholds reached a floor of roughly 0.3 arcmin (18 arcsec) at around
2 – 3 cpd. This is somewhat larger than the 5-10 arcsec best stereo acuity, and may be
attributable to the use of a repeating pattern as the stimulus (resulting in a “wallpaper
illusion”), that limited performance for high spatial frequencies (i.e. resulted in front/back
confusions due to multiple possible planes of fusion). Legge and Gu also found that the
contrast ratio, a difference in contrast between left and right eye stimuli, had a larger effect
on disparity thresholds than overall contrast. This finding is potentially important for
stereoscopic display specification. This is because contrast across the two camera images
may vary, thus producing differences in contrast across the left and right eye images. In fact, this issue has been noted in previous work with remote vision aerial refueling (Kooi & van Breda, 2003). Although they did not report specifically on depth perception, they did note that binocular image luster was quite evident due to differential contrast, and resulted in eyestrain and discomfort. Similarly, differences in contrast sensitivity between the left and right eye might also be expected to affect stereo acuity.

An ongoing debate concerning binocular disparity is the distance over which disparity remains a useful cue to depth. This is because binocular disparity is proportional to the square of the overall distance:

\[ \delta \approx \frac{\Delta D \times IPD}{D^2} \]

Where \( \delta \) is disparity, \( \Delta D \) is the distance between two objects in depth, IPD is interpupillary distance, and D is the viewing distance (Allison, Gillam, & Vecellio, 2009).

Thus, it has generally been assumed that disparity is not useful beyond about 20 feet because monocular cues might be expected to become more informative at larger viewing distances. However, other researchers have disagreed with this assertion (Allison et al., 2009; Patterson & Martin, 1992). Allison et al. (2009) conducted an experiment to examine stereo acuity for viewing distances much larger than typically used in laboratory tests – up to 18 m (60 ft.). In their experiment, Allison et al essentially constructed a supersized Howard depth test (Howard, 1919), where subjects were asked to make front/back discriminations for a large rod mounted on a motorized track relative to a reference panel. Binocular thresholds ranged from approximately 3 – 12 cm, corresponding to stereo acuities of approximately 5 – 10 arcsec. Allison et al. concluded that stereopsis remains
an effective cue for distances larger than typically assumed, and go on to hypothesize that
to hypothesize that for individuals with 5 arcsec thresholds, that it could remain an effective cue for distances
as large as 1 km. Lloyd and Nigus (2012) also concluded that errors in distance estimates
for their boom operator task at a simulated distance of roughly 60 feet were consistent with
stereo thresholds of approximately 10 arcsec.

Clinical Measures of Stereo acuity and Depth Perception

Available Tests

Depth perception and stereo acuity tests can be traced to the WWI and WWII era. These tests include the Howard-Dolman depth perception test (originally described by Howard, 1919); the Optec Vision Tester stereo acuity test, and the AO Vectograph (see Figure 1-8). These tests are all manually administered, and, although appropriate for rapid clinical screening, result in fairly coarse estimates of stereo acuity. For example, the OVT measures stereo acuity only to a level of 15 arcsec. However, for aircrew with good ocular health, stereo acuity thresholds are at the level of 3 to 10 arcsec (Air Force Waiver Guide, 2013; McKee, 1983; Patterson & Martin, 1992). In fact, Howard (1919) actually found that aviators with the best visual function could achieve stereo acuity thresholds as low as 2 arcsec (his Class A observers). Although tests such as the OVT, Titmus, and AO Vectograph are easily and quickly administered, accuracy, test-retest reliability, and efficiency could be improved with the development of a computer-based test. Further, with the introduction of new technologies, such as the stereoscopic remote view system on the new USAF aerial refueling tanker, assessment of ocular alignment and stereovision may become more important, and potentially require a different testing approach.
Bach et al. (2001) demonstrated the feasibility of a computer-based test – the Freiburg Stereo acuity Test, using a relatively low-resolution CRT display and ferroelectric LCD shutter glasses. Two key aspects of the Freiburg Stereo acuity Test are the use of random position offsets to obscure monocular cues, and the use of Gaussian blur/antialiasing to allow sub-pixel changes in position between the left and right eye stereo images. The 800 x 600 pixel display used by Bach et al limited them to a 20-arcsec pixel size at a 4.5 m viewing distance (resulting in a minimum step size of 30 mm in depth for a 65 mm IPD). This would have set a floor for stereo acuity thresholds significantly larger than can be achieved by some observers; however, the use of antialiasing allowed them to adjust each stereo half-image by as little as 1 arcsec (resulting in a minimum step size of
1.5 mm in depth for a 65 mm IPD at 4.5 m). Using this apparatus and procedure, Bach et al showed that thresholds for fine stereopsis as low as 2.5 arcsec could be obtained, which is consistent with previous research (Howard, 1919; McKee, 1983). They also repeated the test with a monocular condition, and for two strabismic patients, which resulted in much higher thresholds. The high monocular thresholds indicated that the steps taken to obscure monocular cues (i.e. slight horizontal offset of the left and right eye stimuli that may be detectable when viewing the stimuli with one eye) were effective. This is potentially a significant issue for vision screening, since some observers may pass the test despite poor ocular alignment, being stereo weak, or even stereo blind (Cooper & Warshowsky, 1977; Fahle, Henke-Fahle, & Harris, 1994; Fawcett & Birch, 2003). For this reason, multiple measures of stereo acuity will be employed in the present experiment.

Although no longer heavily relied upon for routine diagnosis and screening, the Howard-Dolman test, involving mechanically adjusted rods inside a wooden box viewed at 20 feet, is still in use at the USAF for depth perception testing. In the event that aircrew fail both the OVT and the AO Vectograph, they may still be eligible for a waiver if they can perform the HD test with an error of 30 mm or less (roughly equivalent to a 10 arcsec disparity). The HD test is a variant on the original Howard (H) depth test dating to 1919. The HD test is more easily and quickly administered compared to the Howard depth test, which uses the method of constant stimuli rather than direct adjustment by the patient. However, this method permits the patient to use a bracketing technique that may allow them to reduce their error score even though they may actually have defective stereopsis. In fact, Larson (1985) showed that test scores based on the Howard and Howard-Dolman procedures were uncorrelated. Subjects with poor scores on the H test could still achieve
low error scores on the HD test. This is consistent with results observed in practice, patients can fail the OVT and the AO Vectograph (with stereo acuity worse than 60 arcsec), but can still do better than 30 mm on the HD test. Based on his evaluation of the two tests, Larson strongly recommended the H-test over the HD-test, and Westheimer (2013) notes that the Howard test is still one of the best clinical procedures. However, the H test is extremely time consuming. Howard (1919) used the method of constant stimuli, with 10 trials at each of 5 different separation distances of the two rods to obtain a reliable estimate for binocular depth thresholds. The use of a computer-based test and an adaptive staircase procedure would enable obtaining highly accurate thresholds in a time frame comparable to the existing paper-based methods.

Differences in Stereo acuity Test Results

Fawcett and Birch (2003) note that stereo blind patients and patients with strabismus, or who may be recovering from surgery to correct strabismus may often be classified as having some stereo capability after completing the Titmus or Randot tests. This is because these tests contain monocular cues that may allow the patients to identify the targets in the first several rows of the test (i.e. obtain a score of 400 to 140 arcsec). Their own results support this conclusion. Similarly, Cooper and Warshowsky (1977) found that 22% of their 49 test subjects could complete at least 2 lines of the Titmus test, and 4% could accurately complete all 9 lines of the Titmus test using monocular cues alone (i.e. 40 arcsec or better). Thus, some patients may pass existing tests of stereo acuity despite being stereo blind/weak.

Another potential problem for stereo acuity testing is that the prevalence of vergence misalignment may result in a diagnosis of poor stereo acuity, when, in fact, the
observer has good stereo acuity but the presence of vergence misalignment results in placement of the test stimulus well off the horopter. Blakemore (1970) showed that stereo acuity declines exponentially when measured with a depth pedestal of between 0 and 2 degrees. Thus, any vergence misalignment will have a large effect on the measured stereo acuity. Zaroff, Knutelska, and Frumkes (2003) found that although most observers’ highest stereo acuity was obtained at or near the plane of the display (90% within 11 arcmin), several observers’ best stereo acuity was obtained with a pedestal off the plane of the display. Ten percent of their sample had vergence misalignment of greater than 11 arcmin, which is a higher level of prevalence than previously reported. The authors did not believe that undetected phorias could account for these results. These authors also report that the incidence of stereo-blindness in healthy observers less than 60 years of age was approximately 1%, and the incidence of stereo anomaly (confusion between crossed vs. uncrossed disparities) was approximately 8%. These estimates are much lower than other studies have reported (e.g. 20% estimate by Lambooij et al., 2010). They attribute this to excluding subjects due to eye disease and substandard visual acuity and the rigorous technique for measuring stereo acuity that they employed.

Another potentially significant issue for clinical testing of stereo acuity and depth perception is that the old (e.g. OVT, Titmus, AO Vectograph) and new tests (e.g. Freiburg) may have different outcomes. The old tests require only that observers detect something different in one target, while the newer computer-based tests require observers to discriminate between crossed vs. uncrossed disparities (i.e. they have to localize the target in depth in order to respond correctly). Lloyd (2012), for example, found that 8 of 12 subjects were unable to discriminate crossed from uncrossed disparity and were
disqualified from further participation. Similarly, McIntire, Wright, Harrington, Havig, Watamaniuk, and Heft (2014) found that while some observers could pass the Titmus test, they performed poorly at a depth placement task, and were later found to fail a computer-based stereo acuity task requiring depth discrimination. Stereo-anomaly is believed to be fairly common, with estimates as high as 30% of the population with selective deficits for perceiving depth from crossed vs. uncrossed disparity (Wilmer, 2008). However, the prevalence of this proposed deficit may depend on the test methods. For example, the use of short vs. long exposure durations (Patterson & Fox, 1983), or the use of pedestals to control for vergence misalignment (Zaroff et al., 2003). Another possibility, as noted above, is that observers are able to perceive that something is different about the target in some tests and so are able to pass the test without actually being able to localize the target in depth. Fahle et al. (1994) suggest that the combined effects of procedural differences and perceptual learning could result in differences in stereo acuity thresholds as large as a factor of 10 between laboratory and clinical test results (see e.g. Figure 1-9, which shows thresholds obtained using very different procedures and stimuli). Further, some tests may use stimuli consistent with McKee’s (1983) recommendations for best stereo acuity, while others may not. With such a high degree of variability in test methods it is difficult to estimate with any confidence what the prevalence of stereo anomaly actually is.

Utility of Stereopsis

As Wilmer (2008) notes, although much is known about how stereopsis works, little is known about its actual utility. It has been suggested that stereopsis is important for braking camouflage, which may account for why most predators have stereoscopic vision.
Wilmer (2008) calls for research to examine the correlation between stereopsis and various measures of task performance (e.g. depth estimation, table tennis skill, etc.).

![Figure 1-9. Stereo acuity thresholds from Howard (1919) and from Zaroff et al. (2003).](image)

Wilmer goes on to note that “given how little is known about the utility of stereopsis, initial studies might be relatively broad-based and exploratory…”. In aviation, good stereo acuity and depth perception have been viewed as critical to safe flight for many years. Wilmer & Berens (1920) noted that “The value of stereoscopic vision….is of great value in judging distance and landing…The importance of this qualification seems to grow greater as our experience increases”. Howard (1919) developed one of the first tests of depth perception for screening purposes and, on the basis of his research, believed that “to possess normal judgment of distance one’s binocular parallactic angle should not be greater than 8.0” (arcsec)”. However, the debate concerning the utility of depth perception has also been ongoing since the early 1900s. Howard (1919) noted “some examiners have
questioned the absolute necessity of binocular single vision as a preliminary requirement” (for aviation vision standards).

A number of studies have found that good stereo and/or the use of stereo displays is important for athletics (Boden, Rosengren, Martin, & Boden, 2009; Laby et al., 1996), walking (CuQlock-Knopp, Torgerson, Sipes, Bender, & Merritt, 1995; Hayhoe, Gillam, Chajka, & Vecellio, 2009), catching (Mazyn, Lenoir, Montagne, & Savelsbergh, 2004), distance estimation (Lloyd et al., 2012), and telerobotics (Chen, Oden, Kenny, & Merritt, 2010; Held & Hui, 2011). The USAF and US Navy currently require a minimum level of binocular vision and stereo acuity for pilots, and for some non-pilot aircrew (Air Force Instruction 48-123, Medical Examinations and Standards, 2009, Air Force Waiver Guide, 2013, BUMED P-117, 2005; Entzinger, 2009), and the Confederation of Australian Motor Sport concluded in response to a petition from a one-eyed driver that the potential increase in risk was too great, ruling that the driver would not be permitted to compete (Westlake, 2001). Figure 1-10 shows how the current USAF standard for stereo acuity compares to Howard’s (1919) data. Howard’s data are comparable to more recent estimates of stereo acuity thresholds (Zaroff et al., 2003) that show that there is a very wide range of stereo capability. However, the absolute thresholds are much lower for Howard’s data. Figure 1-9 shows the log-transformed stereo acuity thresholds for both Howard (1919, n = 106) and Zaroff et al. (2003, n = 106), and as commented on above, illustrates that substantially different results may be obtained for stereo acuity estimates for different methods and stimuli.
In a recent review by McIntire, Havig, and Geiselman (2012), the utility of stereo displays across a variety of studies was found to be highly variable, with some showing a benefit of stereo, but roughly an equal number showing no benefit.

Figure 1-9. Howard's (1919) stereo acuity data in comparison to current USAF standards.

In another review, Entzinger (2009) concluded that stereo acuity was not relevant for the landing of large aircraft, but could potentially be of some use for the landing of smaller aircraft. Fielder & Moseley (1996) go so far as to note that it was their impression that many ophthalmologists and vision scientists secretly suspect that the utility of having a second eye is simply to have a spare in the event that one is injured. However, another set of studies conducted for the insurance industry in Canada, found that commercial truck drivers with binocular vision problems were more likely to be involved in more severe crashes than normal drivers (Laberge-Nadeau et al., 1996), and taxi drivers with binocular vision problems were more likely to be involved in accidents (Maag, Vanasse, Dionne, & Laberge-Nadeau, 1997). Additionally, numerous studies have demonstrated that contrary to the “spare-eye” hypothesis, individuals who have lost one eye, or have suffered from
amblyopia have significant difficulties with hand-eye coordination tasks, and may have impairments in driving (Grant & Moseley, 2011).

A major limitation for many studies examining the utility of depth perception for performance of real-world tasks is that the measures of depth perception are often coarse and suffer from significant floor effects. If stereo acuity or other clinical metrics relevant to binocular health are actually obtained, they are often limited to, for example, a 40- or 60-arcsec minimum threshold (or in some cases “fly positive,” meaning that subjects could see the 3D fly on a commonly available near stereo acuity test). Since Howard demonstrated that more than 75% of the individuals he tested had depth thresholds better than 10 arcsec (Figures 1-8 and 1-9), it is not surprising that many studies have been unable to clearly demonstrate the utility of depth perception with such limited measures of binocular health. Further, it may also be important to assess depth/stereo under more challenging conditions. A recent study (Yoo, Reichow, & Erickson, 2011) found that even among young, healthy athletes, stereo acuity declined substantially for non-primary gaze (i.e., over-the-shoulder viewing). And, as discussed previously, a significant number of individuals may not actually possess normal stereo acuity despite passing some existing screening tests. Thus, it is not clear that current USAF vision tests and standards will accurately identify individuals for which viewing stereoscopic displays over long periods of time could be problematic.

**Purpose of This Research**

Current USAF vision standards and existing USAF measures of stereo acuity (AFVT, Titmus test, AO Vectograph, and HD test, see discussion and Figure 1-1 above) and ocular alignment (phoria, fusion range) predate the use of stereoscopic displays. Thus,
existing USAF standards and ocular metrics may not be applicable to boom operators using the new KC-46 remote vision system. As noted above, the current standards are designed to uncover only gross deficiencies, and may not identify individuals likely to experience discomfort, or possibly even reduced performance, when using stereoscopic displays for long periods of time. Table 1 below summarizes current USAF vision standards for stereo acuity and ocular alignment (phorias). Aircrew passing the standard summarized in Table 1 are qualified (assuming they also pass all the other medical standards) to fly and do not have to return for additional vision screening unless another visual health issue arises. However, for depth perception (but not phorias), aircrew failing the 25 arcsec minimum score on the AFVT can obtain a waiver if they pass the AO Vectograph at 60 arcsec or better, or, failing that and provided they are not diagnosed with a visual condition such as suppression, can achieve 120 arcsec on the AO Vectograph and 30 mm or better on the HD test. Thus, for depth perception testing in particular, the standard is designed to be inclusive rather than exclusive. With the introduction of the KC-46 and the stereoscopic RVS for aerial refueling, it will be important for aeromedical personnel to be able to identify aerial refueling candidates who may not be medically fit to perform the refueling task, and may pose an unacceptable level of operational risk. Additionally, the ability to accurately identify these individuals during preliminary medical screening could substantially reduce training costs (i.e. reduce the number of training candidates that may wash out). A more refined depth perception standard based more objectively on operationally-relevant performance could also potentially enlarge the pool of eligible candidates through avoiding inadvertently failing suitable candidates.
The goal of this research is to examine the relationship between clinical and laboratory measures of visual performance and ocular alignment and performance on an operational task similar to that expected to be performed by aerial refueling operators (ARO’s) using newly developed remote-view crew stations.

Table 1-1. USAF, USN, and USA Vision Standards. FCIII (scanners) are applicable to aerial refueling operators.

<table>
<thead>
<tr>
<th></th>
<th>USAF (Air Force Waiver Guide)</th>
<th>USN (NAVMED P-117)</th>
<th>USA (Army Regulation 40-501)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereo-acuity (Arc Second)</td>
<td>FCI/IA/II/III (Scanners)</td>
<td>FCII</td>
<td>Class I, Class II (except Fixed Wing Aircrew), Class III (including UAV Operators, Critical Flight Deck Personnel)</td>
</tr>
<tr>
<td>Phoria</td>
<td>Eso</td>
<td>&lt; 10 PD</td>
<td>&lt; 15 PD</td>
</tr>
<tr>
<td></td>
<td>Exo</td>
<td>&lt; 6</td>
<td>&lt; 8</td>
</tr>
<tr>
<td></td>
<td>Hyper</td>
<td>&lt; 1.5</td>
<td>&lt; 2</td>
</tr>
<tr>
<td>Tropia</td>
<td>0</td>
<td>&lt; 15</td>
<td>0</td>
</tr>
</tbody>
</table>

VTA-DP: Vision Test Apparatus-depth perception
OVT: the Optec Vision Tester

The feasibility of using newly developed computer-based measures of depth perception will also be examined. With that goal in mind, several new computer-based tests were identified or developed that could be plausible alternatives to existing clinical tests. These tests included near/far stereo acuity, fusion range, and contrast sensitivity. Although several additional test methods were investigated, because they are either difficult to administer or excessively time consuming, they are not discussed here.
Individual differences in depth perception, ocular alignment, other measures of visual performance, and operationally-relevant performance, as well as ability to tolerate extended viewing periods, were evaluated. The applicability of these vision tests was evaluated using a correlation analysis. Because the stated objective of this research was to evaluate several different vision tests simultaneously, a more conservative level of significance (p < 0.01 rather than p < 0.05) was adopted for the correlation analysis. Additional analyses were performed to evaluate sensitivity and specificity. A practical aspect of this work was to identify a set of vision tests that characterizes an individual’s performance while avoiding redundant or overlapping tests, which are undesirable in the clinic where many applicants (e.g. aerospace medical exams, driver licensing, etc.) must be screened rapidly. Additionally, the effects of extended viewing periods with stereoscopic imagery on performance and comfort are not well known. With the exception of survey results obtained in entertainment venues (e.g. Solimini, 2013), most studies examining discomfort with the use of stereo displays use very short viewing periods (e.g. Kooi & Toet, 2004; Lambooij et al., 2011). However, Shibata et al. (2011) did evaluate reports of discomfort following an approximately 20 minute viewing period under different conditions of VA mismatch. Emoto, Niida, Okano, and Member, 2005 also examined discomfort following 1 hour viewing periods with 3D TV. The proposed study involved approximately 2-hour viewing periods, a duration chosen through in-depth interviews with experienced boom operators, to examine the effects of extended exposure to a stereoscopic display during the simulation of an aerial refueling task.
II. GENERAL METHODS

Apparatus

*Remote Vision System Aerial Refueling Simulation*

A 5-channel PC-based image generator (Flight Safety International Vital X) was used to present the boom model and receiver aircraft model over ground terrain. A standard flight database provided by Flight Safety consisting of desert terrain was used. In this simulation the simulated KC-46 aircraft flies continuously around a “race track” pattern over the western United States. The tanker traveled at approximately 320 kts, at an elevation of 30,000 ft., and completed the circuit in approximately 20 minutes. Two of the IG channels drove the left and right eye views of the simulated boom camera viewpoints. Two Black Magic DVI Extenders and an AJA Video Multiplexer were used to combine the two video channels into one top-and-bottom stereoscopic image which was presented using a ViewSonic V3D231 23-Inch 3D Monitor. Three additional channels drove the left, center, and right panoramic camera viewpoints which were presented on three HP monitors (HP Pavilion 21.5” IPS LED HD monitors). A sixth channel served as the host emulator which allowed the experimenter to control the simulated refueling scenario. The ViewSonic monitor uses passive polarizing glasses to present the left and right eye views. Thus each eye viewed a 1920 x 540 pixel image (approx. 1.1 arcmin/pixel) at 24 Hz when viewing the 3D display. The 3D monitor was mounted in front of the observer and viewed at a distance of 33 inches. The panoramic displays were mounted above the 3D display. Subjects controlled the boom using 2 Saitek X52 flight controllers mounted on either side of the 3D display. The controller on the right provided control over the boom left, right, up and down position. The controller on the left provided control over the boom extension. Light turbulence was simulated using a sum of sines. The relative motion of the two
aircraft occurred in the left-right direction with a peak amplitude of approximately 2 ft. The range of temporal frequencies present in the relative motion was approximately 0.1 to 1 Hz.

An HP Envy 15t Quad laptop computer running the Windows 8.1 OS was also used to control an aspect angle recognition task that was displayed on both the 2D HP monitors and the ViewSonic stereoscopic monitor. For this task, participants entered their responses using a 6-button response box (Cedrus Corp. model RB-730). A Microsoft Surface Pro 2 was used to administer a questionnaire. Figure 2-1 shows the refueling task apparatus.

![Simulated remote vision system aerial refueling task apparatus.](image)

**Figure 2-1.** Simulated remote vision system aerial refueling task apparatus.
Clinical Vision Screening

A Reichert Model No 11636 phoropter was used for the phoria and zone of clear single binocular vision tests. These tests were administered by USAFSAM optometrists (Figure 2-2).

Figure 2-2. Phoropter used to administer phoria and fusion range tests.

Computer-based Vision Screening

A Dell Precision T7610 with Nvidia GeForce GTX 680 graphics card was used to administer a battery of stereo acuity, fusion range, and spatial rotation tests. Participants used either a Logitech game controller or Saitek joystick (Figure 2-3) to enter responses. The test software was developed using Visual Basic (http://msdn.microsoft.com/en-us/library/2x7h1hfk.aspx), Matlab (www.mathworks.com), and Octave (http://www.gnu.org/software/octave/). Tests were displayed on an Asus VG278HE 3D
monitor with 1920 x 1080 pixels that was compatible with Nvidia 3D Vision2 using active shutter glasses. At a 1 m viewing distance the angular pixel size was 1.1 arcmin.

Figure 2-3. Logitech game controller and Saitek joystick used to enter responses for computer-based vision tests.

Contrast Sensitivity Test

A prototype computer-based contrast sensitivity test was provided by Adaptive Sensory Technology. An NEC MultiSync P463 flat panel display was used to present the imagery, which was generated by a Unix-based Shuttle PC. Participants entered responses using a Microsoft X-box game controller.

Stereo-Optical Armed Forces Vision Tester

A Stereo-Optical Armed Forces Vision Tester was used for initial screening to administer the standard visual acuity and stereo acuity vision tests. Responses were recorded manually using paper and pencil score sheets (Figure 2-4).
Figure 2-4. Stereo-Optical Armed Forces Vision Tester.

*Stereo Display Luminance and Crosstalk*

For most of the computer-based vision testing, the participants donned active shutter Nvidia 3D vision glasses. The brightness of the Asus display was noticeably lower in 3D mode compared to 2D mode. In order to increase the luminance, the Nvidia brightness boost was enabled and set to the highest level. Table 2-1 summarizes luminance measurements obtained without glasses for each of the vision tests (described further in General Procedure below).
Table 2-1. Luminance measurements obtained without glasses for the stereo acuity and fusion range tests.

<table>
<thead>
<tr>
<th>Stereo Acuity Test Luminance (cd/m²)</th>
<th>Fusion Range Test Luminance (cd/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Center White Background</td>
<td>81.7</td>
</tr>
<tr>
<td>Gray Background - Center</td>
<td>42.3</td>
</tr>
<tr>
<td>Gray Background - Edges</td>
<td>30.7</td>
</tr>
<tr>
<td>Estimated Contrast</td>
<td>-0.48</td>
</tr>
<tr>
<td>White target (estimated)</td>
<td>81.7</td>
</tr>
<tr>
<td>Gray Background</td>
<td>45.04</td>
</tr>
<tr>
<td>Estimated Contrast</td>
<td>0.81</td>
</tr>
</tbody>
</table>

To evaluate crosstalk, black and white test images were generated using a software application that allowed different images to be displayed to the left and right eyes in 3D view and measured using a Minolta LS-100 through the left and right lenses of the glasses. Crosstalk was calculated according to an equation provided by Weissman and Woods (2011):

\[
OCT_{RL} = (O_{GL} - O_{BL})/(O_{WL} - O_{BL})
\]

Where \( OCT_{RL} \) is observed crosstalk – right to left, \( O_{GL} \) is observed ghost image-left, \( O_{BL} \) is observed black-left, and \( O_{WL} \) is observed white-left. Based on this procedure, a crosstalk level of 0.2% was estimated for the Asus 3D display using the active shutter glasses.

For the aerial refueling task, the participants donned passive, polarized 3D glasses to view the ViewSonic 3D display. The luminance of the F-35 receiver aircraft without glasses was approximately 20.3 cd/m2. The luminance of the receiver aircraft through the glasses was approximately 7.4 fL. Thus the glasses reduce luminance by approximately...
64%. A crosstalk value of 1.4% was obtained based on the same procedure described above. Although no ghosting was apparent when viewing the refueling simulation, this value is greater than the 0.3% recommended by Kooi & Toet (2004) for comfortable viewing. Figure 2-5 shows the two different kinds of 3D glasses used in this research.

![3D glasses](image)

**Figure 2-5.** Nvidia 3D Vision active shutter glasses (left) and passive polarizing 3D glasses supplied with the ViewSonic 3D monitor (right).

**Participants**

A total of 56 participants were administered a set of basic vision tests to screen for different levels of depth perception. Twenty-eight of the 56 participants went on to complete the entire experiment. Of those 28 participants that completed the full experiment 23 were male and 5 were female (83% male, 17% female). Eight participants were over the age of 40 and exhibited varying levels of presbyopia. Two subjects were between the ages of 30 and 40, and 17 subjects were between the ages of 18 and 30. Participants over the age of 40 were included since aerial refueling operators could also be over the age of 40. All 29 of the subjects completed the data collection; however, one subject was
excluded as a malingering. Participants were screened in order to include approximately equal numbers of participants in each of the USAF categories for binocular status (passing, waiverable, and failing – see Table 1-1). Ultimately 14 were recruited that fell into the passing category, 7 in the waiverable category, and 6 in the failing category.

**General Procedure**

This research involved a substantial amount of time for research participants and so, to avoid inducing fatigue, took place over the course of several visits. The initial vision screening took place on day 1, computer-based vision screening took place on day 2, aerial refueling task training took place on day 3, and completion of the 2-hour aerial refueling task took place on day 4. Some participants also returned for a 5th day for experiment 2.

**Vision Screening**

Participants were first screened to determine their binocular vision status according to USAF vision standards and waiver policy (see Table 1-1). All participants were first administered the Stereo-Optical Armed Forces Vision Tester (AFVT). Following this test, USAFSAM Optometrists administered the phoria test, and if necessary, AO Vectograph (AOV) stereo acuity test. The AFVT test was used for stereo acuity test results ranging from 15 to 60 arcsec. For participants failing the AFVT, the AOV was administered, which uses disparities ranging from 60 to 240 arcsec. These test results were combined rather than being reported separately. For participants obtaining a score of 15 to 40 arcsec on the AFVT, that score was used. However, for participants with poorer stereo acuity, where the AO Vectograph was administered, the larger stereo acuity values resulting from that test were used (60 to 240 arcsec). Additionally, the zone of clear single binocular vision was also measured.
Computer-based Vision Tests

Near and far stereo acuity tests

This test involved depth discrimination of a center circle relative to a larger outer circle presented on a stereo-display (Figure 2-6) at a viewing distance of either 1 meter or 4 meters. For this test, the circles were presented for 2 seconds. The participants’ task was simply to indicate whether the small inner circle appeared to be in front of or behind the larger outer circle using the game controller (illustrated in Figure 2-7). For this test, the luminance of the gray background was 38 cd/m² and the luminance of the center disc was 75 cd/m². An adaptive procedure was used to estimate the stereoscopic depth threshold (Psi method; Kingdom & Prins, 2010), with the number of trials fixed at 35. Prior to the actual test, each participant completed several practice trials to become familiar with the test procedure. Auditory feedback (tone: correct; buzzer: incorrect) was provided. This stereo acuity test is similar to a computer-based test described by other researchers (Bach et al., 2001).

Figure 2-6. Near stereo test stimulus.
Figure 2-7. Schematic view of stereo acuity rings. Left: small ring in front of larger ring; center: small ring behind larger ring, right: eye indicating relative viewing position.

Vertical fusion range

This task required that participants indicate when a circle viewed at a distance of 1 meter on the Asus stereo monitor became blurry or doubled ("breaks") using the game controller as the circles moved apart in the vertical direction. The direction of motion then reversed, and the participant next indicated when the circles returned to a single "fused" image using the game controller. This task was repeated several times. The amount of separation between the left and right eye images was recorded at the time the subject pressed the button on the game controller for each trial. The test images viewed by the participant (for left and right eye images) at one instant in time are shown in Figure 2-8.
Figure 2-8. Vertical fusion range stimuli (showing both left and right eye images).

**Horizontal fusion range**

For this test the participants’ task was to indicate when the circle became blurry or doubled (“breaks”) using the game controller as the circles moved apart in the horizontal direction (crossed and uncrossed directions). Next, the direction of motion reversed and the participant indicated when the circles returned to a single “fused” image using the game controller. The separation at which the circles “broke” and then returned to a fused state was recorded. This task was repeated several times. The test images viewed by the participant (for left and right eye images) are shown in Figure 2-9.
Contrast Sensitivity

This test is part of a collaborative effort between USAFSAM and Adaptive Sensory Technology (AST). A specially modified AST contrast sensitivity (CS) test was provided by AST. For this test, participants viewed band-pass filtered Landolt C stimuli at a distance of 4 meters and responded to the orientation (Figure 2-10). The AST CS test uses a sophisticated adaptive procedure to rapidly estimate the contrast sensitivity function (Lesmes, Lu, Baek, & Albright, 2010). Each participant completed the CS test 3 times: 1) binocularly (OU), 2) monocular left eye (OS), and 3) monocular right eye (OD). Diffusing filters were used to equalize the luminance across the two eyes for the monocular conditions. For the binocular condition, neutral density filters were used to maintain the same luminance across all three conditions. No feedback concerning accuracy of responses was provided for this test.
Figure 2-10. Band-pass filtered Landolt C stimuli. Figure provided by AST. Image used with permission.

**Aerial Refueling Task**

The Boeing KC-46 ARO crew station uses two cameras mounted at the rear of the aircraft to generate a stereoscopic view of aircraft approaching the KC-46 for refueling. The imagery is presented to the ARO using a stereoscopic display. Additional cameras (panoramic cameras) and displays allow the ARO to view the airspace around the KC-46 to monitor approaching aircraft and aircraft waiting to refuel. The ARO controls the boom using two joysticks mounted on the console. When a receiver aircraft approaches to within about 1 mile the ARO takes control of the receiver aircraft (i.e. becomes the air traffic controller) and relies primarily on the panoramic cameras to guide the receiver to pre-contact position. At approximately 1000 feet the ARO transitions to the 3D display. The pre-contact position is at approximately 50 feet, at which point the ARO is focused nearly entirely on the stereo display and overlay display (analogous to a heads-up display, or HUD, indicating the boom position and other information). The ARO calls out distances (e.g. 100, 90, 80, 70, …) and issues verbal commands or uses the indicator lights to adjust the position of the receiver aircraft. For heavy aircraft closure rate is about 1 ft/sec, while closure rate may be up to about 5 ft/sec for smaller, fighter aircraft. For large aircraft (e.g. C-17) the off-load time can be 30-40 min, during which time the ARO must monitor the
situation through the RVS. Figure 2-11 shows the view of a KC-10 boom operator during refueling.

![Figure 2-11. View from the KC-10 ARO crew station.](image)

The simulated aerial refueling task for this research was designed based on in-depth interviews with experienced KC-135 and KC-10 boom operators. The simulated refueling task is similar to that of a “fighter drag” refueling scenario where a boom operator repeatedly refuels several aircraft while on a long-haul flight. Prior to beginning the refueling task, each participant also completed a general questionnaire to document eyeglass/contact usage, flight experience, gaming experience, and any previous experience with discomfort when viewing 3D movies. Participants also completed pre- and post-questionnaires concerning discomfort/fatigue. Figure 2-12 shows several of the questionnaire items, which were adapted from Shibata et al. (2011).
Figure 2-12. Questionnaire items adapted from Shibata et al. (2011).

Training

All participants underwent training sessions for the refueling task prior to formal experimental data collection. The training sessions were designed to ensure participants had mastered control of the simulated air refueling boom and could fairly quickly maneuver the boom into contact position before formal data collection. Length of training varied substantially depending on the skill level of individual participants. Thus, training duration ranged from approximately 30 to 120 minutes. All participants also engaged in an additional 10 to 15 minutes of training just before beginning the refueling task for formal data collection. All participants also practiced the aspect angle recognition task during training to make sure they were familiar with the aircraft silhouette for each perspective and with the response entry method.
III. EXPERIMENT ONE – METHOD

Two-hour RVS Aerial Refueling

Experiment 1 was designed to simulate conditions a KC-46 boom operator might experience during an extended “fighter drag” refueling mission. Thus experiment 1 was designed to require participants to switch back and forth between the 2D panoramic and 3D stereoscopic displays and to make repeated refuelings over that time period – conditions that were anticipated to induce visual fatigue or discomfort for some individuals. To simulate the approach phase, an aircraft aspect angle task was employed where each participant was asked to identify the aspect (orientation) of a fighter aircraft viewed through the panoramic displays. Each participant then repeated the aspect task using the 3D display. The Psi method (Kingdom & Prins, 2010) was employed to vary the receiver aircraft position and estimate a threshold recognition range for each participant and for each of six different positions (left panoramic, center panoramic, right panoramic, and left, right, and center in the 3D display). In a given block of trials, participants completed 35 trials while viewing the panoramic display and another 35 trials in the 3D display, which took approximately 5 minutes. Figure 3-1 shows several images of the target aircraft as displayed on the panoramic displays at different distances. Figure 3-2 shows several images of the target aircraft as displayed on the 3D display (i.e. the boom camera view) at different distances. Table 3-1 summarizes the luminance and contrast measurements for the aspect angle aircraft for the two displays and several distances.
Figure 3-1. Images of the aspect angle task target aircraft as displayed on the panoramic displays at different distances.

Figure 3-2. Images of the aspect angle task target aircraft as displayed on the 3D display (boom camera view) at different distances.
**Table 3-1.** Luminance and contrast measurements for the aspect angle aircraft for two displays and several distances.

<table>
<thead>
<tr>
<th>Distance</th>
<th>Aircraft Luminance (cd/m²)</th>
<th>Background Luminance (cd/m²)</th>
<th>Contrast</th>
<th>Distance</th>
<th>Aircraft Luminance (cd/m²)</th>
<th>Background Luminance (cd/m²)</th>
<th>Contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>397</td>
<td>8.2</td>
<td>56.0</td>
<td>-0.85</td>
<td>397</td>
<td>31.5</td>
<td>13.9</td>
<td>1.27</td>
</tr>
<tr>
<td>631</td>
<td>8.4</td>
<td>56.0</td>
<td>-0.85</td>
<td>631</td>
<td>30.4</td>
<td>13.9</td>
<td>1.19</td>
</tr>
<tr>
<td>795</td>
<td>8.1</td>
<td>56.0</td>
<td>-0.86</td>
<td>795</td>
<td>29.1</td>
<td>13.9</td>
<td>1.09</td>
</tr>
<tr>
<td>985</td>
<td>9.2</td>
<td>57.2</td>
<td>-0.84</td>
<td>1001</td>
<td>27.8</td>
<td>13.9</td>
<td>1.00</td>
</tr>
<tr>
<td>1275</td>
<td>9.6</td>
<td>57.1</td>
<td>-0.83</td>
<td>1257</td>
<td>26.7</td>
<td>13.9</td>
<td>0.92</td>
</tr>
<tr>
<td>1598</td>
<td>11.0</td>
<td>57.2</td>
<td>-0.81</td>
<td>1589</td>
<td>25.0</td>
<td>13.9</td>
<td>0.80</td>
</tr>
<tr>
<td>2003</td>
<td>20.1</td>
<td>57.2</td>
<td>-0.65</td>
<td>2003</td>
<td>21.3</td>
<td>13.9</td>
<td>0.53</td>
</tr>
<tr>
<td>2514</td>
<td>23.1</td>
<td>57.2</td>
<td>-0.60</td>
<td>2512</td>
<td>20.8</td>
<td>13.9</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3159</td>
<td>18.0</td>
<td>13.9</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3974</td>
<td>16.0</td>
<td>13.9</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Following completion of the aspect angle task, the refueling task began. At the beginning of the task, the experimenter moved the receiver aircraft into pre-contact position and deployed the boom. The participant was required to recognize when the receiver began moving into the contact position and to then fully deploy the boom and prepare for refueling. Participants were instructed to make contact with the receiver as quickly as possible, while avoiding striking the fuselage with the boom. In the event of a strike, the 3D display flashed to indicate a strike, and the experimenter retracted the boom to indicate a failed trial. The experimenter then repositioned the aircraft into the pre-contact position, then back to contact position and deployed the boom. This repositioning imposed a time penalty. In order to lock the refueling boom to the receiver aircraft’s receptacle, the participant had to successfully maneuver the boom to within 3 inches and depress the trigger button on the right-hand controller. If successful, the boom locked to the receptacle and the participant was instructed to leave the boom in place to simulate fuel off-load for
several seconds. The experimenter then repositioned the receiver to prepare for the next trial. If unsuccessful, the boom would retract, and the participant could try again to make a connection. Following each 7-minute refueling block, the participant completed a short questionnaire to estimate fatigue/discomfort during the course of the experiment. Participants repeated this procedure for 2-hours. The receiver aircraft position was varied by approximately 1 foot fore, aft, up, and down across refueling attempts. Varying the position of the receiver aircraft from trial to trial was intended to prevent participants from simply learning a series of steps to complete the task. Twelve different receiver aircraft refueling positions were used. Figure 3-3 shows four of the 12 receiver aircraft positions. In the pre-contact position (upper left), the receiver aircraft is about 60 ft away from the tanker and out of reach of the boom. In the other three positions, the receiver aircraft is about 40 ft from the tanker. However, as shown, the receiver varies noticeably fore/aft and/or horizontally/vertically by about 1 foot in each direction. The luminance of the receiver aircraft was approximately 20 cd/m². However, when viewed through the 3D glasses, the luminance of the receiver aircraft was reduced to approximately 7.4 cd/m².

The experimental design was intended to require participants to switch back and forth between 2D and 3D displays repeatedly as boom operators will be required to do during a lengthy mission in the KC-46.

Throughout the refueling period, a number of metrics concerning boom and receiver aircraft position were recorded at a 24 Hz sampling rate and recorded to a log file. These metrics included boom (elevation, yaw, extension) and receiver position, nozzle distance to receiver aircraft refueling receptacle, joystick input, and number of strikes. Figure 3-4 shows several seconds of this data output for one subject.
Figure 3-3. Receiver aircraft positions viewed through the simulated RVS.

Pre-contact position (top left), contact position 1 (top right), contact position 4 (lower left), and contact position 5 (lower right).
Figure 3-4. Approximately 30 seconds of data recorded at 24 Hz for one subject. In this example, nozzle distance to receptacle, boom pitch, boom yaw, and boom extension are shown.

Over the course of a 2-hour refueling session, a very large amount of data was collected – approximately 90,000 data points and a file size of roughly 30 MB. These data were reduced to generate the metrics used to evaluate individual refueling performance: Number of contacts, Number of collisions, Number of attempts, Duration, and Distance at attempt. Figure 3-5 shows 80 minutes of recorded data based on these metrics.
Figure 3-5. Boom position data recorded at 24 Hz were reduced to a few metrics that were used to evaluate refueling performance. In this example, time to successful coupling (trial duration), number of attempts (number of trigger presses to initiate boom to receiver lock), and nozzle distance to receptacle at attempt are shown for an 80 minute time period for one subject.

IV. EXPERIMENT ONE – RESULTS AND DISCUSSION

Reported Discomfort

Reports of discomfort generally increased over the course of the 2-hour refueling task. Figure 4-1 shows the average level of eye-tiredness, difficulty maintaining focus/vergence, and viewing comfort throughout the 2-hour task (average value at the end of each 7 minute refueling session). However, although several subjects took one or more rest breaks during the task, no one became so uncomfortable that they quit the task. Although, most subjects only reported mild to moderate discomfort, several subjects
reported being very uncomfortable for a substantial proportion of the task duration. Figure 4-2 shows the number of subjects reporting different levels of eye-tiredness.

**Figure 4-1.** Average values for focus/vergence difficulty, viewing comfort, and eye-tiredness throughout 2-hour refueling task.
Figure 4-2. Histogram showing the number of subjects reporting different levels of eye-tiredness at the conclusion at the end of the 2-hour refueling task.

Standard Vision Test Results

USAFSAM optometrists and optometry technicians administered several standard clinical tests to each subject. These tests included the Armed Forces Vision Tester visual acuity and stereo acuity tests, AO Vectograph stereo acuity test, phoria tests, and zone of clear single binocular vision (ZCSBV). Although ZCSBV is not a standard USAF screening test, it is a standard optometric test. Figure 4-3 shows a histogram for the AFVT and AO Vectograph stereo acuity test results. Based on the results of these two standard USAF stereo acuity tests, subjects generally fall into two categories: 25 arcsec (1.4 log arcsec) or better, or 60 arcsec (1.8 arcsec) and worse. Recall that 25 arcsec or better is required to pass FCIII vision standards, but that stereo acuity as poor as 120 arcsec (2.1 log arcsec) can still be waived. Based on the standard stereo acuity tests alone, 24 of the 27 subjects either pass the standard, or fall into the waiverable category.

The range of near and far horizontal phoria measures is shown in Figure 4-4. Note that exophorias are represented with negative values. As shown, subjects tend to shift from more esphoric at far, to more exophoric at near. Far vertical phorias were also measured for each subject. Two subjects had vertical phorias of 1 PD and one subject had a vertical phoria of 2 PD. The rest of the subjects had vertical phorias of 0 PD.
Figure 4-3. AFVT/AO Vectograph stereo acuity frequency histogram.

Figure 4-4. Frequency histograms for horizontal far (top) and near (bottom) phorias.
The results of the ZCSBV test is shown for 2 subjects in Figure 4-5. Break and recovery points for base-in and base-out across 3 accommodative distances are shown. Subject 33 (left) has a relatively narrow ZCSBV, while subject 139 (right) has a relatively wide ZCSBV. The ZCSBV was reduced to a single value by estimating the area between the base-in and base-out recovery points. A frequency histogram for ZCSBV area is shown in Figure 4-6.

**Figure 4-5.** ZCSBV for two subjects.

**Figure 4-6.** ZCSBV area frequency histogram.
Experimental Vision Test Results

Near and far stereo acuity

Near and far stereo acuity test result frequency histograms are shown in Figure 4-7. As shown, the computer-based stereo acuity test results differ substantially from the USAF standard stereo acuity test results. Figure 4-8 shows the relationship between USAF standard stereo acuity tests and experimental computer-based stereo acuity test. The two test methods are significantly correlated ($r = 0.62, p < 0.001$). Because the AFVT stereo test and AO Vectograph are both far stereo acuity tests, the correlation with the far computer-based stereo acuity test is shown.

![Near Stereo Acuity Histogram](image1)

![Far Stereo Acuity Histogram](image2)

**Figure 4-7.** Near and far stereo acuity histograms.
Figure 4-8. Relationship between USAF standard stereo acuity tests and experimental computer-based stereo acuity test.

Figure 4-9 shows the relationship between the experimental near and far stereo acuity tests. The near and far stereo acuity tests were significantly correlated (r = 0.79, p < 0.0001). Because these two tests were highly correlated, the average of the near and far stereo acuity thresholds for each participant were used in subsequent analysis rather than analyzing them separately.

Figure 4-9. Relationship between experimental near and far stereo acuity tests.
**Fusion Range**

Figure 4-10 shows the log transformed horizontal and vertical fusion ranges for two subjects. Subject 87 has a relatively small horizontal fusion range while subject 76 has a relatively larger fusion range. Figure 4-11 shows the relationship between horizontal fusion range (recovery) and ZCSBV. The two tests were not significantly correlated ($r = 0.38$, $p = 0.05$). In subsequent analyses, log horizontal and vertical fusion range values were combined (log vertical range subtracted from log horizontal). Horizontal and vertical fusion ranges were combined in this way based on discussions with USAFSAM optometrists and ophthalmologists. A **larger** horizontal range is indicative of good binocular alignment and ocular motility. Conversely, a **smaller** vertical range is indicative of good ocular alignment.

**Figure 4-10.** Horizontal and vertical fusion ranges for two subjects. Fusion break points are shown in blue, fusion recovery points are shown in red (log arc minutes).
Figure 4-11. Relationship between log horizontal fusion range and log ZCSBV area.

Figure 4-12. Contrast sensitivity (CS) test results for two subjects. CS functions are shown for binocular (OU), left eye only (OS), and right eye only (OD).

Contrast Sensitivity

Figure 4-12 shows contrast sensitivity (CS) test results for two subjects. Subject 87 has a relatively low contrast sensitivity, and has lower contrast sensitivity in the right eye compared to the left eye. Subject 6 has relatively greater contrast sensitivity and less difference in contrast sensitivity between the two eyes. The area under the log contrast
sensitivity function (AULCSF) was used as the metric to compare against refueling performance in subsequent analyses. A frequency histogram showing the range of AULCSF scores (OU) is shown in Figure 4-13.

![Frequency histogram showing the range of AULCSF scores (OU)](image)

**Figure 4-13.** Frequency histogram showing the range of AULCSF scores (OU).

**Correlations Among Vision Tests**

Table 4-1 shows the correlations among all the vision tests administered to participants in Experiment 1. Significant correlations (p < 0.01) are highlighted in green. In general, the computer-based tests are correlated with each other. However, a notable exception is that stereo acuity was not correlated with contrast sensitivity or acuity. Among the standard tests, the AFVT/AOV was correlated with the experimental computer-based stereo acuity test, but also fusion range. Vertical phoria was also correlated with fusion range, and near and far phoria were correlated.
Table 4-1. Correlations among vision tests.

<table>
<thead>
<tr>
<th></th>
<th>OVT-AOV Stereo</th>
<th>H Phoria (far)</th>
<th>V Phoria (far)</th>
<th>H Phoria (near)</th>
<th>ZCSBV</th>
<th>Stereo-acuity</th>
<th>Fusion Range</th>
<th>CS (OU)</th>
<th>CS (Min)</th>
<th>Acuity (OU)</th>
<th>Acuity (Min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OVT-AOV Stereo</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H Phoria (far)</td>
<td>0.07</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V Phoria (far)</td>
<td>0.35</td>
<td>-0.19</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H Phoria (near)</td>
<td>-0.09</td>
<td>0.73</td>
<td>-0.36</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZCSBV</td>
<td>-0.40</td>
<td>-0.07</td>
<td>-0.05</td>
<td>0.26</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stereo-acuity</td>
<td>0.67</td>
<td>-0.06</td>
<td>0.38</td>
<td>-0.35</td>
<td>-0.47</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fusion Range</td>
<td>-0.55</td>
<td>0.00</td>
<td>-0.53</td>
<td>0.29</td>
<td>0.46</td>
<td>-0.72</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS (OU)</td>
<td>-0.31</td>
<td>-0.17</td>
<td>-0.42</td>
<td>0.15</td>
<td>0.15</td>
<td>-0.36</td>
<td>0.60</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS (Min)</td>
<td>-0.19</td>
<td>-0.17</td>
<td>-0.40</td>
<td>0.11</td>
<td>0.14</td>
<td>-0.35</td>
<td>0.65</td>
<td>0.85</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acuity (OU)</td>
<td>-0.30</td>
<td>-0.17</td>
<td>-0.25</td>
<td>0.08</td>
<td>0.14</td>
<td>-0.18</td>
<td>0.46</td>
<td>0.80</td>
<td>0.55</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Acuity (Min)</td>
<td>-0.27</td>
<td>-0.29</td>
<td>-0.40</td>
<td>0.03</td>
<td>0.13</td>
<td>-0.31</td>
<td>0.58</td>
<td>0.84</td>
<td>0.89</td>
<td>0.71</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Aircraft Aspect Angle Recognition Range

Figure 4-14 shows average aircraft aspect angle recognition range for each viewing condition. Many of the subjects had difficulty correctly identifying aircraft aspect for the left versus right panoramic camera views, which is reflected in the decreased recognition range for those viewing conditions.

![Figure 4-14](image.png)

Figure 4-14. Average aircraft aspect angle recognition range for each viewing condition.
The relationship between aspect angle recognition range for the center panoramic viewing condition and visual acuity is shown in Figure 4-15 \((r = 0.44, p = 0.02)\).

**Figure 4-15.** Relationship between AST visual acuity and aspect angle (AA) recognition range for the center panoramic display.

**Aerial Refueling Performance**

To generate a single aerial refueling metric, each of the individual metrics were \(z\)-transformed and summed to generate an overall measure of RVS refueling performance according to the following equation:

\[
AR\ \text{Performance} = N\ \text{Contacts} Z - N\ \text{Collisions} Z - N\ \text{Attempts} Z - \text{Mean Duration} Z - \text{Mean Distance} Z
\]

Two of the metrics (number of contacts and duration) were sensitive to the efficiency of the experimenter when manually controlling the receiver aircraft position and boom deployment. The first subjects to complete the experiment tended to complete fewer connections on average, while subjects completing the experiment at a later date tended to achieve a higher number of connections. This was because each different task was initiated by the experimenter, who became more efficient over time (e.g. switching back and forth
between control panels on the IG). This had a small but cumulative effect on time-dependent measures. Thus, prior to calculating the combined aerial refueling metric, the values for Number of Contacts and average duration were adjusted for this bias. This was accomplished by plotting each metric according to order of participation, fitting a curve, then normalizing the values to obtain a slope of zero. Figure 4-16 shows the range of values obtained for Number of Contacts per block, Number of Collisions per block, Number of Attempts (average number misses prior to a successful connection), mean duration to make contact, mean distance across attempts, and the composite aerial refueling performance score (AR Performance, based on combined z-scores).

**Standard USAF Vision Test relationship to Aerial Refueling Performance**

Figure 4-17 shows the relationship between four of the USAF standard vision tests and AR performance. The correlations between the USAF standard stereo tests ($r = 0.295$, $p = 0.135$), far phoria ($r = 0.006$, $p = 0.98$), near phoria ($p = 0.45$, $p = 0.019$), and AR Performance were not significant. However, the correlation between vertical phoria and AR Performance was significant ($r = 0.73$, $p < 0.001$). The horizontal dashed lines in Figure 4-17 indicate 1 standard deviation below the mean for aerial refueling performance and the vertical solid lines indicate the current FCIII minimum vision standard for each test vision test. Using one standard deviation below the mean for refueling performance is an arbitrary decision, but indicative of clearly below average performance (analogous to receiving a grade of “D” or less on a test graded on a curve).
Figure 4-16. The range of values obtained for Number of Contacts per block (top left), Number of Collisions per block (top right), average Number of Attempts (middle left), Mean Duration (middle right), Mean Distance (bottom left), and composite aerial refueling (AR) performance (bottom right).
Based on these criteria, the sensitivity and specificity of the FCIII standard tests were also computed:

\[
Sensitivity = \frac{\#True Positives}{\#True Positives + \#False Negatives}
\]

\[
Specificity = \frac{\#True Negatives}{\#True Negatives + \#False Positives}
\]

These values can also be defined as the probability that a test is positive given that the patient is ill, or the probability that the test is negative given that the patient is well, respectively. Table 4-2 summarizes sensitivity and specificity values for each of the standard tests as well as for participants classified according to the USAF FCIII vision standard. For the purposes of this comparison, participants falling into the waiver category are considered as passing. According to these criteria, true negatives, true positives, false negatives, and false positives can be identified based on which quadrant each score falls into on each plot. For example, in Figure 4-17 (upper left) for AFVT/AOV stereo test results, scores falling into the lower right quadrant represent participants correctly identified as failing (true positives); scores falling into the upper right quadrant represent participants incorrectly identified as failing (false positives); upper left quadrant as correctly identified as passing (true negatives), and lower left quadrant as incorrectly identified as passing (false negatives).
Figure 4-17. The relationship between OVT/AOV stereo (top left), far phoria (top right), near phoria (bottom left), and vertical phoria (bottom right) and composite AR performance. The horizontal dashed lines indicate 1 standard deviation below the mean for aerial refueling performance and the vertical solid lines indicate the current FCIII minimum vision standard for each test vision test.

Table 4-2. Sensitivity and specificity values for each of the standard USAF vision tests.

<table>
<thead>
<tr>
<th>Standard USAF Tests</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCIII</td>
<td>1.00</td>
<td>0.84</td>
</tr>
<tr>
<td>OVT/AOV</td>
<td>1.00</td>
<td>0.96</td>
</tr>
<tr>
<td>Far Phoria</td>
<td>0.50</td>
<td>0.92</td>
</tr>
<tr>
<td>Near Phoria</td>
<td>1.00</td>
<td>0.80</td>
</tr>
<tr>
<td>Vertical Phoria</td>
<td>0.50</td>
<td>0.92</td>
</tr>
</tbody>
</table>
Figure 4-18 shows the relationship between the same four USAF standard vision tests and average discomfort ratings. The correlation between the USAF standard stereo tests and average discomfort was not significant ($r = 0.19$, $p = 0.34$). The correlations between far phoria, near phoria, and average discomfort were not significant ($r = 0.06$, $p = 0.78$; $r = 0.43$, $p = 0.02$). However, the correlation between far vertical phoria and average discomfort was significant ($r = 0.73$, $p < 0.001$).

**Figure 4-18.** The relationship between AFVT/AOV stereo (top left), far phoria (top right), near phoria (bottom left), and vertical phoria (bottom right) and average discomfort ratings.

**FCIII Status and Aerial Refueling Performance**

Figure 4-19 shows the relationship between FCIII status and AR Performance. An ANOVA reveals that the effect of FCIII classification for AR Performance was not significant [$F(2,24) = 3.05$, $p = 0.07$]. Figure 4-19 also shows 1 standard deviation below
the mean for AR Performance (dashed line). Based on this criterion, the sensitivity of the FCIII classification is 1.0 and the specificity is 0.84. As shown, several individuals identified as failing the FCIII standard perform at or above average AR Performance (i.e. the test results in some false positives).

Figure 4-19. The relationship between FCIII status and AR performance.

One standard deviation below the mean for AR success rate is indicated by the dashed line.

ZCSBV and Aerial Refueling Performance

Figure 4-20 shows the relationship between ZCSBV, AR Performance, and average discomfort. ZCSBV was not correlated with AR Performance or average discomfort ratings ($r = 0.12, p = 0.56; r = 0.01, p = 0.97$). Figure 4-20 (left) also illustrates the sensitivity and specificity for the ZCSBV (0.0 and 0.88). The horizontal dashed line indicates 1 standard deviation below the mean for AR Performance and the vertical solid line indicates 1 standard deviation below the mean for ZCSBV.
Figure 4-20. The relationship between log ZCSBV and AR Performance (left), and to average discomfort (right).

Experimental Vision Tests and Aerial Refueling Performance

Figure 4-21 shows the relationship between average stereo acuity, fusion range and AR Performance. The correlation between stereo acuity and refueling performance was significant ($r = 0.56$, $p = 0.002$). The correlation between fusion range and AR Performance was also significant ($r = 0.70$, $p < 0.001$). One standard deviation below the mean is shown in Figure 4-21 for AR Performance (dashed horizontal line) and each vision metric (solid vertical lines). For the stereo test, sensitivity is 1.0 and specificity is 0.88. For fusion range, sensitivity is 1.0 and specificity is 0.92.

Similarly, Figure 4-22 shows the relationship between average stereo acuity, fusion range and average discomfort ratings. The correlation between fusion range and average discomfort ratings was significant ($r = 0.55$, $p = 0.003$), but the correlation between stereo acuity and discomfort was not significant ($r = 0.4$, $p < 0.04$).
Figure 4-21. The relationship between average stereo acuity (left), fusion range (right) and AR Performance. One standard deviation below the mean is shown for AR Performance (dashed horizontal line) and each vision metric (solid vertical lines).

Figure 4-22. The relationship between average stereo acuity (left), fusion range (right) and average discomfort ratings.

**Contrast Sensitivity and Aerial Refueling Performance**

Figure 4-23 shows the relationship between contrast sensitivity (OU) and minimum contrast sensitivity (i.e. the contrast sensitivity of the weakest eye) and refueling performance. The correlations between binocular contrast sensitivity, minimum contrast sensitivity, and aerial refueling performance are both highly significant ($r = 0.58$, $p = 0.002$;
$r = 0.68$, $p < 0.001$). The dashed lines indicate one standard deviation below the mean for refueling performance. The vertical solid lines indicate one standard deviation below the mean for contrast sensitivity (OU) and for minimum contrast sensitivity. Based on these criteria, the sensitivity of both the binocular contrast sensitivity and minimum contrast sensitivity tests is 1.0 and 0.92.

**Figure 4-23.** The relationship between contrast sensitivity (left), minimum contrast sensitivity (right) and refueling performance.

**Figure 4-24.** The relationship between binocular contrast sensitivity (left) and minimum contrast sensitivity (right) and average discomfort ratings.

Similarly, Figure 4-24 shows the relationship between contrast sensitivity (OU) and minimum contrast sensitivity (i.e. the contrast sensitivity of the weakest eye) and average discomfort ratings. The correlation between binocular contrast sensitivity and average
discomfort was not significant \( (r = 0.35, p = 0.08) \), nor was the correlation between minimum contrast sensitivity and average discomfort \( (r = 0.39, p = 0.04) \).

**Effect of Age on Discomfort**

Figure 4-25 shows the average discomfort ratings for two age groups: 30 years of age and younger and over 30 years of age. A one-tailed t-test shows that the effect of age on discomfort ratings was significant \( [t(25) = 2.05, p = 0.03] \).

![Figure 4-25. Difference in discomfort ratings by age group.](image)

Figure 4-26 shows the relationship between minimum contrast sensitivity and average discomfort and between minimum contrast sensitivity and aerial refueling performance for participants 30 and younger. The correlation between minimum contrast sensitivity and average discomfort when participants over 30 are excluded was highly significant \( (r = 0.73, p < 0.001) \), as was the correlation between minimum contrast sensitivity and aerial refueling performance \( (r = 0.8, p < 0.001) \).
Figure 4-26. Relationship between minimum contrast sensitivity and average discomfort (left) and between minimum contrast sensitivity and aerial refueling performance for participants 30 and younger.

For participants over 30, the relationship between minimum contrast sensitivity and performance was significant ($r = 0.84, p = 0.003$) but not for discomfort ($r = 0.06, p = 0.87$) as shown in Figure 4-27.

Figure 4-27. Relationship between minimum contrast sensitivity, discomfort ratings, and refueling performance for participants over 30 years of age.

Similarly, the relationship between near phoria and average discomfort was greater for participants 30 and younger (Figure 4-28). The correlation between near phoria and
average discomfort was significant ($r = 0.64$, $p = 0.006$) as was the correlation between near phoria and aerial refueling performance ($r = 0.71$, $p = 0.001$).

![Graphs showing the relationship between near phoria and discomfort/performance](image)

**Figure 4-28.** The relationship between near phoria and average discomfort (left) and between near phoria and aerial refueling performance (right) for participants 30 and younger.

However, near phoria does not appear to be predictive of either discomfort ($r = 0.17$, $p = 0.64$) or performance ($r = 0.1$, $p = 0.79$) for participants over the age of 30 as shown in Figure 4-29.

![Graphs showing the relationship between near phoria, discomfort, and performance](image)

**Figure 4-29.** Relationship between near phoria, discomfort, and refueling performance for participants over the age of 30.
Discomfort and Refueling Performance

Figure 4-30 shows the relationship between average discomfort ratings and aerial refueling performance for participants over 30 (left) and participants under 30 (right). Increasing discomfort was highly correlated with decreased refueling performance for younger participants ($r = 0.89$, $p < 0.001$), but not for older participants ($r = 0.19$, $p = 0.6$). As shown in both figures 4-30 and 4-25, older participants generally reported lower levels of discomfort.

![Image of Figure 4-30 showing the relationship between average discomfort ratings and aerial refueling performance for participants over 30 (left) and participants under 30 (right). Increasing discomfort was highly correlated with decreased refueling performance for younger participants ($r = 0.89$, $p < 0.001$), but not for older participants ($r = 0.19$, $p = 0.6$). As shown in both figures 4-30 and 4-25, older participants generally reported lower levels of discomfort.

Figure 4-30. The relationship between average discomfort ratings and aerial refueling performance. Results for participants over 30 years of age is shown on the left, results for participants 30 and younger is shown on the right.

Combined Vision Scores and Refueling Performance

The three computer-based vision test scores (minimum contrast sensitivity, fusion range, and stereo acuity) were combined to produce a combined vision metric (CVM). The CVM is the sum of the z-transformed values for each test. Figure 4-31 shows the relationship between CVM and aerial refueling performance. The correlation between CVM and aerial refueling was highly significant ($r = 0.76$, $p < 0.001$). Adopting a criterion
of 1 standard deviation below the mean for CVM resulted in a sensitivity of 1.0 and a specificity of 0.92.

Figure 4-31. The relationship between the combined vision metric (CVM) and aerial refueling performance. One standard deviation below the mean aerial refueling performance is indicated with the horizontal dashed line. The solid vertical line represents 1 standard deviation below the CVM mean. The dashed vertical line indicates 2 standard deviations below the CVM mean.

Summary of Experiment 1 Results

Tables 4-3 and 4-4 summarize the correlations between each vision test and simulated remote vision system aerial refueling performance as well as the sensitivity/specificity results for each test. Table 4-5 summarizes the correlations between each vision test and average discomfort ratings. Tests with correlations significant at p < 0.01 and sensitivity and specificity greater than 0.9 are highlighted in green.
Table 4-3. Correlations between USAF standard vision tests and AR performance (left), and between experimental vision tests and AR performance (right). Tests resulting in a significant correlation (p < 0.01) are highlighted in green.

<table>
<thead>
<tr>
<th>Standard USAF Tests</th>
<th>Correlation</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCIII</td>
<td>0.34</td>
<td>0.080</td>
</tr>
<tr>
<td>OVT/AOV</td>
<td>0.29</td>
<td>0.140</td>
</tr>
<tr>
<td>Far Phoria</td>
<td>0.00</td>
<td>0.980</td>
</tr>
<tr>
<td>Near Phoria (all)</td>
<td>0.45</td>
<td>0.020</td>
</tr>
<tr>
<td>Near Phoria (&lt;30)</td>
<td>0.71</td>
<td>0.002</td>
</tr>
<tr>
<td>Vertical Phoria</td>
<td>0.73</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experimental Tests</th>
<th>Correlation</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereo</td>
<td>0.56</td>
<td>0.002</td>
</tr>
<tr>
<td>Min CS (all)</td>
<td>0.68</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Min CS (&lt;30)</td>
<td>0.80</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Fusion Range</td>
<td>0.70</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CVM</td>
<td>0.77</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ZCSBV</td>
<td>0.16</td>
<td>0.422</td>
</tr>
</tbody>
</table>

Table 4-4. Sensitivity and specificity for USAF standard vision tests and experimental vision tests (right). Tests resulting in a high level of both sensitivity and specificity (> 0.9) are highlighted in green.

<table>
<thead>
<tr>
<th>Standard USAF Tests</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCIII</td>
<td>1.00</td>
<td>0.84</td>
</tr>
<tr>
<td>OVT/AOV</td>
<td>1.00</td>
<td>0.96</td>
</tr>
<tr>
<td>Far Phoria</td>
<td>0.50</td>
<td>0.92</td>
</tr>
<tr>
<td>Near Phoria</td>
<td>1.00</td>
<td>0.80</td>
</tr>
<tr>
<td>Vertical Phoria</td>
<td>0.50</td>
<td>0.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experimental Tests</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereo</td>
<td>1.00</td>
<td>0.88</td>
</tr>
<tr>
<td>Min CS</td>
<td>1.00</td>
<td>0.92</td>
</tr>
<tr>
<td>Fusion Range</td>
<td>1.00</td>
<td>0.92</td>
</tr>
<tr>
<td>CVM</td>
<td>1.00</td>
<td>0.92</td>
</tr>
<tr>
<td>ZCSBV</td>
<td>0.00</td>
<td>0.88</td>
</tr>
</tbody>
</table>
Table 4-5. Correlations between each vision test and average discomfort ratings.

<table>
<thead>
<tr>
<th>Standard USAF Tests</th>
<th>Correlation</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCIII</td>
<td>0.300</td>
<td>0.128</td>
</tr>
<tr>
<td>OVT/AOV</td>
<td>0.192</td>
<td>0.338</td>
</tr>
<tr>
<td>Far Phoria</td>
<td>0.056</td>
<td>0.781</td>
</tr>
<tr>
<td>Near Phoria (all)</td>
<td>0.432</td>
<td>0.024</td>
</tr>
<tr>
<td>Near Phoria (&lt;30)</td>
<td>0.638</td>
<td>0.006</td>
</tr>
<tr>
<td>Vertical Phoria</td>
<td>0.726</td>
<td>0.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experimental Tests</th>
<th>Correlation</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereo</td>
<td>0.399</td>
<td>0.039</td>
</tr>
<tr>
<td>Min CS (all)</td>
<td>0.399</td>
<td>0.039</td>
</tr>
<tr>
<td>Min CS (&lt;30)</td>
<td>0.729</td>
<td>0.000</td>
</tr>
<tr>
<td>Fusion Range</td>
<td>0.550</td>
<td>0.003</td>
</tr>
<tr>
<td>CVM</td>
<td>0.530</td>
<td>0.004</td>
</tr>
<tr>
<td>ZCSBV</td>
<td>0.002</td>
<td>0.993</td>
</tr>
</tbody>
</table>
V. EXPERIMENT TWO – EFFECT OF STEREEO ON REFUELING PERFORMANCE

Participants

Thirteen participants ranging in age from 18 to 57 were recruited to participate in Experiment 2. All thirteen participants also completed Experiment 1 and were highly practiced. Nine of the participants in Experiment 2 were male, and 4 were female. Two experienced KC-135 boom operators also completed the refueling task but did not participate in any of the vision testing.

Apparatus

The same apparatus described in section II was used in Experiment 2.

Procedure

For Experiment two, participants engaged in the same aerial refueling task described in section II for Experiment 1. However, several changes to the procedure were made. In this experiment, the viewing conditions switched between hyper-stereoscopic (same as Experiment 1), normal stereo (camera separation set to an average interpupillary distance of 65 mm), and 2D. The experimenter toggled between hyper- and normal-stereo using a camera configuration setting in the Vital X IG control panel, and could switch between stereo and normal 2D using the AJA video multiplexer control panel. The order of viewing conditions was randomized for each participant. Additionally, for Experiment 2 participants were not asked to engage in the aspect angle recognition task. Thus, participants viewed the central display for nearly the entire duration of the experiment. The refueling duration was also reduced to 5 minutes. Following each 5 minute refueling period, the experimenter changed the viewing condition according to the randomized order.
The same aerial refueling metrics described for Experiment 1 were used for Experiment 2. However, in order to compare performance across viewing conditions, a new metric, aerial refueling success rate, was used. This metric was defined according to the following equation: \( \text{Success Rate} = \frac{\# \text{Contacts}}{\text{Time}} \) (i.e. average number of successful contacts per minute).

All participants had previously engaged in a period of training (for the hyper-stereo viewing condition) prior to beginning Experiment 2. Most of the Experiment 2 participants had already completed Experiment 1 prior to beginning Experiment 2. However a few participants completed Experiment 2 prior to completing Experiment 1. Prior to beginning Experiment 2, participants engaged in 10-30 minutes of additional practice. Participants then completed 3 viewing conditions x 4 repetitions = 12 blocks of aerial refueling. However, the first block for each viewing condition was excluded from subsequent analysis to reduce practice effects.
VI. EXPERIMENT TWO – RESULTS

Figure 6-1 shows the difference in aerial refueling success rate for each viewing condition. Paired sample t-tests revealed that the difference in aerial refueling performance between the 2D and Normal viewing conditions were significant [t(14) = -3.6, p = 0.003], between the 2D and Hyper viewing conditions were significant [t(14) = -5.1, p < 0.001], and between the Normal and Hyper viewing conditions were significant [t(14) = -2.4, p = 0.03].

![Figure 6-1. Number of contacts/minute for each viewing condition.](image)

Although the purpose of Experiment 2 was to investigate the overall effect of viewing condition, the effect of quality of vision was also examined. Figure 6-2 shows the difference in aerial refueling performance for participants categorized as either “good” or “poor” based on the combined vision metric described in Experiment 1. Because there were fewer participants in Experiment 2, participants were simply categorized as above or below the mean (i.e. positive vs. negative combined z-scores). According to these criteria, 7 participants were categorized as “good” and 6 as “poor”. As shown, participants with
better vision appear to benefit from the use of hyper-stereo while participants classified as poor according to the CVM appear to perform worse in the hyper-stereo viewing condition, and an analysis of variance revealed a significant interaction \( F(2,22) = 4.9, p = 0.036 \). These results suggest that some individuals may have more difficulty with stereo viewing conditions relative to normal 2D displays, or at least may not benefit from the use of hyper-stereo.

![Figure 6-2](image_url)

**Figure 6-2.** Aerial refueling performance for participants categorized as either “good” or “poor” based on their combined quality of vision scores.

Figure 6-3 shows aerial refueling performance for each participant in each viewing condition. Participants with combined vision metric scores above the mean are color coded in green while participants with combined vision metric scores below the mean are color coded in red. The results are highly variable, however, participants classified as “good” generally increase or maintain performance across viewing condition, while participants
classified as “poor” generally decline in performance across viewing conditions, particularly for the hyper-stereo condition. Two exceptions to this trend include participant 110, classified as “good” and participant 129, classified as “poor”. Participant 110 did score poorly on the stereo acuity test, which may account for this participant’s relatively poorer performance using stereo displays. However, none of the vision tests account for participant 129’s relatively good aerial refueling performance. The two boom operators performance is shown for comparison, however, since they did not complete the vision testing, thus their combined vision metrics are not available.
Figure 6-3. Aerial refueling performance for each participant in each viewing condition. Participants with combined vision metric scores above the mean are color coded in green while participants with combined vision metric scores below the mean are color coded in red.
VII. GENERAL RESULTS AND DISCUSSION

This research investigated the applicability of current USAF Flying Class III (FCIII) vision standards and vision test methods to performance and viewing discomfort for simulated remote vision system aerial refueling. The simulated remote vision system used in this research effort was designed to be very similar to the remote vision system that will be used by aerial refueling operators selected for the new USAF tanker, the KC-46. The simulation was designed based on specifications provided by the KC-46 Program Office and Boeing. In a recent review paper Winterbottom et al. (2014) note that it is not surprising that research often fails to find a significant relationship between vision test results and task or performance, or fails to find clear effects of the use of stereoscopic displays, because many of the vision tests, and stereo acuity tests in particular, are often coarse. Thus, rather than rely solely on current standard tests to assess quality of vision, this research also examined the applicability of several newly developed vision tests to RVS aerial refueling performance. The general results of each experiment and the implications they may have in relation to previous research are described below. Each section below tries to answer a specific hypothesis or major question of interest, which had initially guided the formulation of the two experiments.

Do USAF FCIII vision standards and tests predict RVS aerial refueling performance?

Because the FCIII vision standard is a combination of multiple tests, the composite and each component are addressed separately.

**FCIII Classification**

No. As summarized in Tables 4-3 and 4-5, FCIII classification was not predictive of either simulated aerial refueling performance or reported discomfort when using the
simulated RVS for a period of 2-hours. However, it is important to note that despite the lack of correlation between FCIII standards and AR performance, the current FCIII standard does achieve a sensitivity of 1.0, indicating that the worst performers are probably being screened out by existing test methods. However, the specificity of the current standard was fairly poor (0.84), indicating that current methods are likely to fail some individuals that may perform the RVS refueling task quite well. However, it is important to note that the 1 standard deviation below the mean criteria is arbitrary, and the sensitivity/specificity analyses presented here are really dependent on a relatively few individuals that clearly fail each of the tests. A much larger number of subjects will be required in future research to adequately address this issue before recommendations concerning passing/failing criteria can be made.

*Standard AFVT/AOV Stereo Tests*

No. The AFVT/AOV stereo test results in particular, which are a significant component of the FCIII classification, were not predictive of RVS AR performance (r = 0.29, p = 0.13). As shown in Figure 4-8, participants obtaining a passing score of 15 arcsec on the AFVT may obtain a test result of anywhere from as low as 5 arcsec (0.74 log arc sec) to as high as 224 arcsec (2.35 log arc sec). Thus, these tests, with only a limited number of trials, and only crossed disparities, may not accurately assess stereo acuity.

*Far Phoria Test*

No. The far phoria test results, another major component of FCIII classification, were not predictive of performance (r = 0.006, p = 0.98). For this research, participants with a wide range of ocular health and visual capability were recruited. With regard to phoria, test results ranged from -25 PD, a very severe exophoria to +9 PD, esophoria falling
outside the normal range. Thus the lack of a significant correlation cannot be attributed to inadequate sampling/range restriction.

**Near Phoria**

Tentatively, yes. For all participants, the correlation between near phoria and AR performance approached significance ($r = 0.45$, $p = 0.02$). The phoria test results reported here were based on a single measurement for each participant. It is possible that the phoria test procedure could be improved, reducing the test-retest variability, and may be a more effective predictor of both AR performance and comfort. Further, for younger, non-presbyopic, participants, near phoria was significantly correlated with AR performance and comfort ratings ($r = 0.71$, $p = 0.002$; $r = 0.64$, $p = 0.006$). Thus, near phoria, but not far phoria, may be applicable to the RVS aerial refueling task. Young participants with large exophoria at near were more likely to report higher levels of discomfort. The correlation between exophoria at near and discomfort may reflect the fact that young participants who tend to under-converge at near distances, may have difficulty maintaining an over-converged (hyper-stereoscopic) ocular posture, particularly for long periods of time. The sensitivity of this test, at 1.0, was good, but low specificity (0.8) suggests that the occurrence of false positives may be an issue for this test.

**Far Vertical Phoria**

Yes. Far vertical phoria, which also factors into FCIII classification, was predictive of RVS AR performance ($r = 0.73$, $p < 0.001$). However, the sensitivity of this test (0.5) was fairly poor. Near phoria was not predictive of AR performance overall, but was predictive of performance for young participants ($r = 0.71$, $p = 0.002$). As with near phoria, it is possible that the phoria test procedure could be improved, reducing the test variability,
and could be a more effective predictor of both AR performance and comfort. As shown in Figure 4-17, vertical phoria test results are restricted to 3 values: 0, 1, or 2 prism diopters. Refining the test to obtain a more continuous test result could potentially improve its ability to predict RVS AR performance.

While this research shows that the current FCIII vision standard is not predictive of simulated RVS aerial refueling performance, the current standard does nonetheless seem to accurately identify those individuals likely to perform particularly badly. Additionally, two of the tests, vertical phoria and near phoria (at least for younger aircrew) may be particularly applicable as screening tests for KC-46 ARO candidates. One additional factor that may not be adequately captured in this research is also important to note. Experienced medical personnel administering the current test battery may not rely entirely on the results of a single test, and often investigate further with additional tests, or alterations to existing tests (e.g. repeating the AOV at a different orientation). So, ARO candidates that may pose a potential operational risk are very likely to be identified as a result of current screening methods. However, as noted above, potentially suitable ARO candidates are probably also being screened out.

The FCIII vision standard and screening procedure is quite similar to the criteria described by Lambooij et al. (2011). They report that classifying individuals on the basis of tests of phoria, fusion range (analogous to the ZCSBV test described here), and several other standard optometric measures is predictive of reading performance and discomfort when using a 3D display. Thus, their research provides additional evidence that individuals with poor/marginal binocular status are likely to experience degraded performance and increased discomfort with the use of 3D displays, and that optometric tests can potentially
be used to identify individuals that may be at risk of degraded performance/increased discomfort. However, whereas the classification procedure described by Lambooij et al. (2011) successfully identifies individuals with poorer performance and increased discomfort in their reading task, the FCIII standard, based on the current results, appears to be less effective. There are several differences which may account for the differing results. First, the classification procedure described by Lambooij et al does not incorporate a stereo acuity test, which is an important aspect of the FCIII standard (and which was not correlated with RVS refueling performance/discomfort). Second, their procedure incorporated some additional tests, such as fixation disparity and fusion amplitude, which are not part of the FCIII standard. Finally, the 3D reading task employed by Lambooij et al (2011) involved very large disparities (up to 1.5 deg) on a display viewed at a very near distance (0.4 m). This configuration was designed to be very taxing. While Lambooij et al show that this configuration does effectively predict group performance and discomfort ratings on the 3D reading task for individuals with marginal vs. good binocular status, the visual conditions in the present research were designed to reflect the operating conditions that will be found with the USAF KC-46 RVS. The RVS is clearly not intended to be uncomfortable for most viewers, and so the current research was attempting to predict much more subtle visual deficits and degradations in performance and comfort. Thus, while the current FCIII standard tests were not particularly effective for predicting more subtle deficits associated with the RVS, it is quite likely that the current standard could effectively identify individuals with reduced performance performing a very challenging 3D display reading task with a 1.5 degree disparity.
Was Zone of Clear Single Binocular Vision (ZCSBV) predictive of RVS AR performance?

No. The correlation between ZCSBV and AR performance and discomfort ratings was near zero. Based on previous research (Shibata et al., 2011), ZCSBV was anticipated to be predictive of discomfort for this task involving a hyper-stereoscopic display system. The use of this display system requires some degree of vergence-accommodation mismatch, which Shibata et al. demonstrate is a source of discomfort. There are a number of differences between the current research and the work by Shibata et al. First, although large disparities are possible with the hyper-stereoscopic RVS, for most of the duration of the refueling task the VA mismatch may fall within the zone of comfort for many observers. While it is very apparent that properly viewing the hyper-stereoscopic display requires adjustment of one’s ocular posture, the position of the receiver aircraft, and, in particular, the position of the refueling receptacle, is located just behind the convergence point of the cameras for much of the viewing period. However, the fact that age has a significant effect on reported discomfort suggests that VA mismatch may still be a significant source of discomfort. Older participants, with varying degrees of presbyopia, and therefore with reduced accommodative range (i.e. regularly experiencing some degree of VA mismatch, to which they have to adapt), reported much lower levels of discomfort in comparison to younger observers. Second, the ZCSBV measurement procedure used in the present research was based on a single measurement, whereas Shibata et al. averaged over 3 repeated measurements. Third, Shibata et al. derived multiple predictors to apply to a variety of different viewing conditions, whereas a single value representing the total area of the ZCSBV was derived for each subject in the current research. Although this was
expected to be a more robust measure (analogous to the area under the contrast sensitivity function), it may not be effective for predicting discomfort for a particular accommodative and vergence distance. Finally, the RVS simulation also introduced another potential source of discomfort distinct from VA mismatch – vertical misalignment, or dipvergence. Thus, increased discomfort scores could be attributable to lack of tolerance to dipvergence rather than VA mismatch, or the combination of the two. This added source of discomfort may account for why vertical phoria, and the experimental fusion range test, which incorporated vertical fusion range, are highly predictive of both discomfort and AR performance. Vertical misalignment can induce significant discomfort quickly. Kooi & Toet (2004) reported that vertical misalignment of 2 PD (1.14 degrees) resulted in the highest levels of discomfort relative to several other stereo image manipulations. Vertical misalignment exceeding 1 degree is a very large vertical misalignment, but it is still notable that this manipulation induced significant discomfort after a viewing period of just 5 seconds. Other research has also demonstrated that smaller amounts of vertical misalignment lead to discomfort, and recommended tolerances vary widely. However, an industry accepted tolerance is approximately 5 arcmin (Melzer & Moffitt, 1997). Outside the center of the RVS stereoscopic display, the vertical misalignment exceeds this value, and could have been a significant factor contributing to discomfort.

**Were the newly developed computer-based vision tests predictive of RVS aerial refueling performance?**

Because several tests were evaluated, each test is again addressed separately.
Stereo Acuity

Yes. An improved, computer-based, stereo acuity test was predictive of simulated RVS refueling performance ($r = 0.56$, $p = 0.002$). This is a much different result than was obtained using the standard AFVT/AVD stereo acuity test, which was not predictive of RVS AR performance ($r = 0.29$, $p = 0.14$). Thus, this research supports the recommendation by Winterbottom et al. (2014) that future research examining the importance of stereo vision and/or stereo displays must also take into consideration the adequacy of the screening test. Hsu, Pizlo, Chelberg, Babbs, and Delp (1996) have similarly noted that there are large individual differences in stereo acuity and recommend administering stereo acuity tests prior to examining the utility of stereo displays. As noted in the introduction, several other researchers have also questioned the utility of some existing tests of stereo acuity (Bach et al., 2001; Cooper & Warshowsky, 1977; Fawcett & Birch, 2003; Larson, 1985). As McIntire, et al (2014) note, their research “may be the first to report that for viewers with clinically normal stereopsis, there is a strong significant relationship between stereo acuity and performance on an S3-D virtual object precision placement task.” That relationship was uncovered only because they used a threshold-level stereo acuity test (also developed by USAFSAM) very similar to the one employed in the current research. Although the current standard tests are correlated with the threshold-level stereo acuity test, as shown in Figure 4-2, there is clearly a substantial floor effect and a large amount of variability in stereo acuity that simply isn’t captured by the existing standard tests. However, stereo acuity test methods require additional research. The threshold level stereo acuity test developed for this research is clearly a dramatic improvement over existing standard tests. But, test-retest reliability could potentially be
improved. The threshold estimate error metric generated by the Psi method for the stereo acuity test resulted in an average error of approximately 0.3 log units, which is fairly large. Although it is possible that stereo acuity measures are inherently noisy, providing stronger accommodative targets to encourage participants to maintain accommodation at the plane of the 3D display could potentially improve the reliability of this test.

**Fusion Range**

Yes. The computer-based fusion range test developed as part of this research effort was predictive of RVS AR performance ($r = 0.70, p < 0.001$) as well as discomfort ratings ($r = 0.55, p = 0.003$). This test combined horizontal fusion and vertical fusion range. The fact that this test was predictive of both performance and comfort suggests that it is important to measure not just horizontal fusion range, but vertical fusion range as well. As noted previously, vertical fusion range may be particularly relevant for displays where vertical misalignment may be present. McIntire et al. (2014) also found that fusion range was predictive of performance on a virtual object placement task with a stereoscopic 3D display. The procedure described by McIntire et al (2014) to measure fusion range was very similar to the method described here, and suggests that a computer-based fusion range test should be incorporated into future research concerning the use of stereoscopic displays.

**Is the viewing distance for the vision test important?**

Potentially, yes. One question relevant to developing a platform specific vision standard for the KC-46 (or other career fields involving the use of stereo displays viewed at a near distance) is whether the vision test should be administered at near vs. far viewing distances. Based on discussions with USAFSAM ophthalmologists and optometrists, far vision tests are given more weight, and deemed more relevant to most aviation tasks, since
those tasks have historically involved the ability to see objects at long distances. However, for the KC-46 ARO RVS, the displays are all viewed at a near distance, and so the importance placed on far vision tests may need to be re-evaluated. The fact that near phoria is predictive of RVS refueling performance while far phoria is not, even for younger participants, suggests that the distance at which the vision test is administered may be an issue deserving of further study. Additionally, with the exception of the far stereo test and contrast sensitivity, the experimental vision testing was done at a near viewing distance (1 m, which is very similar to the RVS viewing distance). It is possible, for example, that the correlation between the experimental fusion range test and AR Performance was strengthened because it was administered at a near viewing distance similar to that of the viewing distance for the RVS. However, as noted in Experiment 1, there does not appear to be a substantial difference between near vs. far stereo acuity in terms of correlation with RVS refueling performance viewed at a near distance (r = 0.52, p = 0.005; r = 0.54, p = 0.004). Additionally, despite some variability in the stereo threshold estimate, near and far stereo tests are highly correlated (r = 0.79, p < 0.001). Thus, viewing distance does not appear to have a large effect on stereo acuity thresholds.

**Was discomfort related to RVS aerial refueling performance?**

Yes. Reported discomfort was highly correlated with refueling performance. Participants reporting high levels of discomfort tended to have substantially worse performance than participants reporting low levels of discomfort. Fortunately, a substantial proportion of the participants in this research reported only relatively low levels of discomfort/fatigue, indicating that widespread reports of discomfort for the fielded system in the future are unlikely. However, these results also suggest that any reports of significant
discomfort should be taken seriously since they may be indicative of degraded performance.

**Was contrast sensitivity relevant to RVS AR performance?**

Yes. The results of this study suggest that minimum (weakest eye) contrast sensitivity may be the single best predictor of RVS aerial refueling performance. And, for younger participants, minimum contrast sensitivity may be the single best predictor of visual discomfort. Although previous research (Kooi & Toet, 2004) has provided some evidence that visual acuity in the weakest eye could be related 3D viewing discomfort, these results were somewhat unexpected. Past research indicated that overall reduced contrast did not have a large effect on minimum disparity thresholds. So long as targets are visible, subjects are generally able to detect even small disparities. However, differing contrast across the two eyes had a much larger effect on disparity thresholds than overall contrast (Legge & Gu, 1989). Legge and Gu (1989) also showed that contrast sensitivity was correlated with disparity threshold at the same/similar spatial frequency ($r = 0.84$). Part of the reason a CS test was incorporated into this research effort was that difference in CS between the eyes was anticipated to be predictive of stereo acuity. However, contrast sensitivity was not found to be correlated with stereo acuity. In the current research, area under the log contrast sensitivity function was used rather than contrast sensitivity at individual spatial frequencies. Although this is likely a very robust measure of overall quality of vision, the lack of a significant correlation between contrast sensitivity and stereo acuity could be partly attributable to this difference in measurement technique. Additionally, Legge and Gu were comparing grating disparity stimuli directly to grating contrast sensitivity, very similar stimuli, whereas the stimuli in the current study involved
pairs of rings and band-pass filtered letters. Although they are somewhat similar, they are not as directly equivalent as the grating stimuli used by Legge and Gu. Figure 7-1 shows the relationship between contrast sensitivity, minimum contrast sensitivity, visual acuity, minimum visual acuity, and stereo acuity. As shown, the relationship is in the right direction (i.e. an inverse relationship, where an increasing CS/VA score tends to be associated with a decreasing stereo acuity threshold), although none of the correlations are significant. However, binocular and minimum contrast sensitivity are approaching significance.

![Figure 7-1](image)

**Figure 7-1.** The relationship between contrast sensitivity (upper left), minimum contrast sensitivity (lower left), visual acuity (upper right), minimum visual acuity (lower right), and stereo acuity.

Kooi and Toet (2004) found that minimum visual acuity was negatively correlated with stereo acuity ($r = -0.96$). They also found that good visual acuity was correlated with
increased visual discomfort, particularly for image misalignment distortions. This finding is essentially the opposite of the results reported here. In the present study, reduced contrast sensitivity, especially in the weakest eye, was correlated with increased discomfort. Kooi and Toet used a similar number of subjects (24) with a similar age range (18-58). Their vision test methods were comparable to the OVT VA and stereo acuity tests. However, whereas individuals with poorer quality of vision were specifically recruited for the present study, those individuals were specifically excluded from the research conducted by Kooi and Toet (stereo acuity of at least 60 arcsec). The TNO stereo acuity test employed in their research uses random dot stereograms, which Fawcett and Birch (2003) argue should be less susceptible to monocular cues that may allow stereo-weak or stereo-blind individuals to obtain passing scores. However, it is possible that accurately perceiving the small dots composing the random dot stereogram requires better visual acuity than the low-pass filtered rings used in the computer-based stereo test in the present study. For the purposes of USAF vision screening, the lack of correlation between visual acuity and stereo acuity may be useful since that implies that the two tests are measuring different aspects of vision.

**VIII. CONCLUSIONS AND FUTURE WORK**

Based on the research presented here, several conclusions can be drawn concerning quality of vision and stereoscopic remote vision system aerial refueling performance and vision screening methods. First, quality of vision is clearly an important human systems factor to consider for the use of stereoscopic remote vision systems such as the USAF KC-46 will use in the near future, and that several other Air Forces have already fielded (e.g. Japan, Italy, Australia and the Netherlands). The use of hyper-stereo clearly improved aerial refueling performance for some observers – those with good quality
of vision. However, the results also show that individuals with poorer quality of vision experience not only increased levels of discomfort but degraded performance that could pose an operational risk if selected to be an aerial refueling operator, or at a minimum, could increase training costs if adequate screening methods are not in place. These findings may also have implications for the selection of operators using stereoscopic displays in surgical applications, or for future remotely piloted systems that may adopt stereoscopic displays. While vision standards have been in place for many decades for aviators, there are no vision standards in place for surgeons using newly developed stereoscopic surgical systems. Although many researchers have looked at discomfort with the use of commercially available 3D TV, these results may also provide some additional insight on which viewers may be particularly susceptible to discomfort in entertainment venues.

Second, this research demonstrates that current vision standards and screening methods were generally not very effective for predicting individual refueling performance or viewing discomfort. One of the primary objectives of this research was to investigate the applicability of current USAF vision standards for medical screening of Flying Class III aircrew that may be selected to operate the KC-46 remote vision system for aerial refueling. Although two of the vision tests that make up the FCIII standard were correlated with simulated RVS refueling performance, FCIII classification was not predictive of either performance or discomfort. However, it is important to note that existing standards do identify the worst performers from among the participants in this research. Additionally, military personnel responsible for medical screening often engage in additional testing if they believe certain individuals may have passed certain tests but may nonetheless have visual deficits that could be problematic. Thus it is unlikely that
individuals posing significant operational risk will slip through the system. A more likely drawback of the current standard may be that potentially suitable candidates are removed from consideration after failing a test that may not be predictive of operational performance.

Although existing standard tests were not highly effective, several experimental tests were identified that were more effective at predicting RVS refueling performance and/or discomfort. Minimum contrast sensitivity was the single best predictor of RVS refueling performance and was also highly correlated with reported discomfort for young (non-presbyopic) participants. A major drawback to contrast sensitivity testing in the past has been the time-consuming nature of the test. However, the test employed in the current research was designed to be rapidly administered and can be completed in about 5 minutes. Thus a complete test, assessing CS for each eye, can be completed in about 10 minutes, which is a reasonable time frame for a busy clinic responsible for screening large number of patients, especially after considering additional benefits concerning automation and electronic medical records that reduces transcription time.

Two additional experimental tests, stereo acuity and fusion range, were demonstrated to predict RVS refueling performance, and fusion range was also predictive of reported discomfort. The fusion range test is rapidly and easily administered, and is a plausible test to be implemented in a clinical setting. The fact that a computer-based threshold level stereo acuity test was correlated with refueling performance while a standard test of stereo acuity was not suggests that careful consideration should be given to the type of stereo acuity test employed in future research concerning the utility of stereopsis and the use of stereo displays. The combination of minimum contrast sensitivity,
stereo acuity, and fusion range tests resulted in a high correlation with aerial refueling performance, and, in comparison to any of the standard measures, resulted in greatly improved sensitivity and specificity (i.e. accurately identifying not only poor performers, but also good performers).

**Recommendations for Future Research**

Although several experienced boom operators and engineers associated with the KC-46 evaluated and approved of the fidelity of the RVS simulation employed in this research, there were several limitations to the current RVS simulation that should be addressed in future research concerning the applicability of vision screening methods for RVS refueling performance.

*Improvements to RVS Simulation*

First, the current simulation lacked sophisticated collision detection. The lack of this feature required a simplification to the refueling procedure – rather than guiding the boom to the slipway surrounding the refueling receptacle and allowing the fuel nozzle to automatically lock in place, the participant was required to manually lock the boom in place. While this is still realistic, since boom operators engage in a similar procedure when the auto-lock fails, it is not the normal procedure. The lack of collision detection also prevented obtaining a measurement of severity of impact. Although the current research examined the number of collisions, which was certainly a valid measure of performance, severity of collision could add an important dimension to the assessment of RVS refueling performance. Additionally, although the refueling simulation used in this research incorporated turbulence in the form of pseudo-random horizontal motion of the receiver aircraft, the addition of turbulence to disturb the receiver aircraft altitude, thus introducing
more variation in stereoscopic depth, would be an improvement on the current procedure. One other limitation of the current simulation is that no instrumentation was provided (e.g. a boom position display). While the focus on visually-guided boom control was intentional, future research should incorporate instrumentation for two reasons. First, instrumentation could affect performance (e.g. help users to calibrate position/distance or, alternatively, distract users from the primary task). Second, the instrumentation introduces an additional depth plane (similar to a HUD), that could also affect performance and comfort.

**Experimental Conditions**

Several recommendations can be made concerning experimental viewing conditions that should be investigated in future research. The RVS simulation used in the current research replicated viewing conditions for a perfectly configured system (perfectly aligned cameras, equal contrast across cameras, etc.). Future research should examine the effect of degraded viewing conditions on user performance and discomfort, which is a very real possibility (Kooi & van Breda, 2003). It is very likely that individuals with poor or even marginal quality of vision could experience substantially degraded performance under these conditions. Additionally, future research should more thoroughly examine the effect of different RVS configurations. The current research showed that performance improved with the introduction of stereo, and further improved with the introduction of hyper-stereo. However, performance with hyper-stereo is clearly dependent on quality of vision. As remote vision systems are deployed more widely, it will be important for system engineers, human factors researchers, and medical personnel to understand and anticipate how performance and medical standards may be impacted with different types of technologies.
Specifically, the relative advantages and disadvantages of orthostereoscopic, hyperstereoscopic, and 2D remote vision systems should be examined.
**VIV. REFERENCES**

*Air Force Instruction 48-123, Medical Examinations and Standards.* (2009).

*Air Force Waiver Guide.* (2013). Wright-Patterson AFB, OH.


